Downloaded from orbit.dtu.dk on: May 04, 2023

DTU Library

# Order statistics and multivariate discrete phase-type distributions.

**Campillo Navarro, Azucena**

*Publication date:*
2019

*Document Version*
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

*Citation (APA):*
Campillo Navarro, A. (2019). *Order statistics and multivariate discrete phase-type distributions.* DTU Compute. DTU Compute PHD-2018 Vol. 492

**General rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Order Statistics and Multivariate discrete phase-type distributions

Azucena Campillo Navarro

Kongens Lyngby 2018
PhD-2018-492

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

PhD-2018-492
ISSN 0909-3192

# Summary (English)

In this thesis, we present three different topics of research which are related to the theory of phase-type distributions. Those topics are explained next.

The first research work is on order statistics from matrix-geometric distributions in the case of a sample of independent and non-identically distributed random variables. We prove that order statistics from matrix-geometric distributions are matrix-geometric distributed and we provide representations for their distributions.

The second research work is a study of the discrete version of multivariate phase-type distributions introduced by V. G. Kulkarni. We give an expression for the joint probability-generating function in the similar way than in the continuous time case and under this base we make an analysis of this class of distributions and present examples that are commonly found in the literature.

The third research work presented came out with the aim of relating the last two topics. That is, we found a problem which relates the concept of order statistics and multivariate phase-type distributions introduced by V. G. Kulkarni, the last in the case of continuous time. Thus, we present a research on concomitants of phase-type distributions. We provide a procedure to calculate the density function of concomitants of phase-type distributions and we prove that concomitants of phase-type distributions are phase-type distributed.

# Summary (Danish)

I denne afhandling præsenterer vi tre forskellige emner inden for forskning, der er relateret til teorien om fase-type fordelinger.

Den første del af forskningsarbejdet omhandler ordensstatistikker fra matrix-geometriske fordelinger for situationer med en stikprøve af uafhængige og ikke-identificerede tilfældige variabler. Vi beviser, at ordensstatistikker fra matrix-geometriske fordelinger er matrix-geometrisk fordelt, og vi giver repræsentationer af deres fordelinger.

I den anden del af forskningsarbejdet studerer vi den diskrete version af multivariate fasefordelinger introduceret af V. G. Kulkarni. Vi udtrykker den fælles sandsynligheds-genererende funktion på samme måde som for den kontinuerte tid, og på basis af dette foretager vi en analyse af denne klasse af fordelinger samt de nuværende eksempler på denne fordeling, som findes i litteraturen.

Den tredje del af forskningsarbejdet i denne afhandling har til formål at sammenholde de to forrige emner. Det vil sige, at vi fandt et problem, der vedrører begrebet ordensstatistik og multivariate fasefordelinger introduceret af V. G. Kulkarni, hvoraf den sidstnævnte er for kontinuert tid. Derfor præsenterer vi en undersøgelse af fasetype-fordelingernes concomitanter. Vi tilvejebringer en procedure til beregning af tætheds-funktionen af fase-typefordelingernes concomitanter, og vi beviser, at concomitanter af fasetypefordelinger er fase-type fordelt.

# Preface

The present thesis is a research work that was carried out at the Section of Statistics and Data Analysis of the Department of Applied Mathematics and Computer Science (DTU Compute) at the Technical University of Denmark and it has been submitted in partial fulfilment of the requirements for acquiring the Danish Ph.D. degree.

The thesis is based on the theory of phase-type distributions and its generalisation called matrix-exponential distributions, in both discrete and continuous time. The contribution work is mainly for multivariate phase-type distributions in the discrete time and for the distribution of concomitants in the continuous time. The focus of this thesis is to present theoretical results that helps to unify the theory of this area.

The thesis consists on three research works: representations of order statistics from matrix-geometric distributions, the discrete version of multivariate phase-type distributions introduced by V. G. Kulkarni and distributions of conconmitants of phase-type distributions.

Lyngby, 14-October-2018

Azucena Campillo Navarro

# Acknowledgements

I would like to start with my supervisor Mogens Bladt. I want to give you many thanks to you Mogens for have been an inspiration to me to work with the theory of phase-type distributions, for your guidance in the research, the enormous effort that you put on me, for having me introduced to this gorgeous country and also for bringing to my life a lot of nice experiences. I could easily continue with many more thanks to you, but I sadly think I have no way to pay you back all you did to contribute in my formation as better person and researcher. I REALLY appreciate it and THANK you.

I also need to thank Mogens again for having me introduced an excellent professor at DTU, my supervisor Bo Friis Nielsen. I want to thank you Bo mainly for being with me tremendously patient since the beginning of my research work, also for having created an efficient dynamic of work and motivated to me to be better at work.

Thanks to my colleagues at the section. All of you were really nice with me. I enjoyed all the lunch times, cake and coffee breaks and some parties that we spent together. Certainly that made my Phd studies more fun and pleasant. Specially many thanks to my officemates Sofie, Agnes and Oscar. You were with me since the beginning of the PhD until the end. I feel very lucky for having your friendship and for having shared my daily PhD life with you.

Finally, I want to express my deep sense of gratitude to my family and friends in Mexico. I thank you for cheering me inexhaustibly all the time. You absolutely encourage me. I love you.

# Contents

CHAPTER 1

# Introduction

## Overview of Phase-type distributions and Matrix-exponential distributions, both discrete and continuous time.

In stochastic modelling, the work related to the class of phase-type distributions started with A. K. Erlang's work [Erl17], where he modelled a lifetime random variable as a finite sum of independent and identical exponentially distributed times. After that work, a relation between the method used by A. K. Erlang and a finite-state Markov jump process was noticed by A. Jensen [Jen53], where he considered a finite sum of independent and non-identical exponentially distributed times. Later, M. F. Neuts [Neu75] showed the tractability of the algebraic formalism of Erlang's approach by generalising Erlang's idea with the so-called method of phases (or stages).

The method of phases adapts an exponential time with fixed rate as the time between the transitions of a Markov jump process (which corresponds to one phase), so the concatenation of $p$ exponential times with different rates is adapted to be the time until absorption of a Markov jump process with $p$ transient states. This technique showed to be an effective approach for studying stochastic systems which can be embedded into Markov jump processes and to develop algorithmically tractable solutions in closed matrix forms, such as formulas for densities, Laplace transforms and moments.

The resulting distribution provided by M. F. Neuts is defined as the distribution of the time until absorption of a particular type of Markov jump process.

In the continuous setting, consider a Markov jump process $\{X_t\}_{t\geq 0}$ with finite state space $\{1, 2, \ldots, p, p+1\}$, where states $\{1, 2, \ldots, p\}$ are transient and the state $p+1$ is absorbing. The density function of the time until absorption of $\{X_t\}_{t\geq 0}$ is given by

$$\boldsymbol{\alpha} e^{\mathbf{S}x}\mathbf{s}, \quad x > 0,$$

where $\boldsymbol{\alpha}$ corresponds to the vector of initial probabilities of the transient states of $\{X_t\}_{t\geq 0}$, $\mathbf{S}$ is the matrix of intensities between transient states and $\mathbf{s} = -\mathbf{S}\mathbf{e}$, where $\mathbf{e}$ denotes the $p$-dimensional column vector of ones. The distribution of the time until absorption of $\{X_t\}_{t\geq 0}$ is named Phase-type distribution.

For the discrete time version of phase-type distributions, consider a Markov chain $\{Y_n\}_{n\geq 0}$ with finite state space $\{1, 2, \ldots, p, p+1\}$, where states $\{1, 2, \ldots, p\}$ are transient and the state $p+1$ is absorbing. The probability mass function of the time until absorption of $\{Y_n\}_{n\geq 0}$ is given by

$$\boldsymbol{\pi}\mathbf{T}^{m-1}\mathbf{t}, \quad m \in \mathbb{N},$$

where $\boldsymbol{\pi}$ is the vector of initial probabilities of the transient states of $\{Y_n\}_{n\geq 0}$, $\mathbf{T}$ is the matrix of transition probabilities between transient states and $\mathbf{t} = \mathbf{e} - \mathbf{T}\mathbf{e}$, where $\mathbf{e}$ is the $p$-dimensional column vector of ones. The distribution of the time until absorption of $\{Y_n\}_{n\geq 0}$ is called discrete phase-type distribution.

Phase-type distributions, in the case of continuous and discrete time, enclose a large number of important distributions, for example the Exponential distribution, Erlang distribution, Geometric distribution and Negative binomial distribution. Further, as a class of distributions it has relevant properties. One important property is that the class of phase-type distributions is dense in the set of distributions defined on the nonnegative real line, this means that any distribution defined on the nonnegative real line can be approximated arbitrary close by a phase-type distribution (weak convergence). Furthermore, this class is closed under finite convolutions, finite mixtures and order statistics.

Another important generalisation of Erlang's work [Erl17] was made by Cox in [Cox55], where he considered complex-valued transitions between the states, which immediately led to the class of distributions with rational Laplace-Stieltjes transform. That class coincides with the class of distributions with density function of the form

$$\boldsymbol{\beta} e^{\mathbf{B}x}\mathbf{b}, \quad x > 0,$$

where $\boldsymbol{\beta}$, $\mathbf{B}$ and $\mathbf{b}$ can possibly have complex entries (see [AB95]) and it is referred as Matrix-exponential distributions.

The discrete time version of the Matrix-exponential distribution is called Matrix-geometric distributions and, similarly, it is defined as the class of distributions with probability mass function of the form

$$\boldsymbol{\gamma}\mathbf{G}^{m-1}\mathbf{g}, \quad m \in \mathbb{N},$$

or, equivalently, the class of distributions with rational probability-generating function.

In Chapter 2 and 3 we present the most relevant elements of the theory of Matrix-geometric distributions and Matrix-exponential distribution, respectively, for the understanding of this thesis.

The class of Matrix-exponential distributions (or Matrix-geometric distributions) contain the class of distributions of phase-type (or discrete phase-type), but the probability interpretation provided by the underlying Markov jump process is not longer valid.

Matrix-exponential distributions share the same closure properties than the class of phase-type distributions in the continuous and in the discrete time. That means that Matrix-exponential distributions are also closed under finite convolutions and finite mixtures; nevertheless, regarding order statistics, only in the continuous time it has been shown that Matrix-exponential distributions are closed under this operation in [BN17, p. 234]. Therefore, in Chapter 4, we prove that Matrix-geometric distributions are also closed under order statistics and we provide representations for their distributions.

## Overview of multivariate phase-type distributions in discrete and continuous time.

Many multivariate stochastic models have been proposed for modelling dependent random variables in many areas. In the theory of phase-type distributions (especially in the continuous time) it has also been considered different concepts for the multivariate case of these distributions. The first class of multivariate phase-type distributions was introduced by D. Assaf in [ALSS84] where it is defined a multivariate phase-type distribution by considering a partition of the state space of an underlying Markov jump process and every marginal is defined as the total time the underlying Markov jump process spends in one of the subsets of the partition before it gets absorbed. In this way, marginals are phase-type distributed, the multivariate class is mathematically tractable and shares to have the same form of the formulas for densities, Laplace transform and moments than the univariate case.

Another relevant class of multivariate phase-type distributions was introduced by V. G. Kulkarni in [Kul89]. In that paper V. G. Kulkarni stated that in the univariate case a phase-type distributed random variable can be seen as the total accumulated reward of a Markov jump process until it gets absorbed. We explain this concept more detailed next.

Consider a Markov jump process $\{X_t\}_{t\geq 0}$ with finite state space $\{1, 2, \ldots, p, p+1\}$, where states $\{1, 2, \ldots, p\}$ are transient and the state $p+1$ is absorbing. The matrix

$$\begin{pmatrix} \mathbf{S} & -\mathbf{Se} \\ \mathbf{0} & 0 \end{pmatrix}$$

is the intensity matrix, where $\mathbf{S}$ is a $p$-square nonsingular matrix, and $(\boldsymbol{\alpha}, \alpha_{p+1})$ denotes the vector of initial probabilities of $\{X_t\}_{t\geq 0}$. Consider

$$\tau = \inf\{t \geq 0 : X_t = p+1\}$$

this is $\tau$ is the time until absorption of $\{X_t\}_{t\geq 0}$.

Let $\mathbf{r} = (r(1), \ldots, r(p))^\top$ be a nonegative $p$-dimensional column vector, where $r(i)$ is interpreted as the reward obtained when the Markov jump process $\{X_t\}_{t\geq 0}$ is in state $i$. We call $\mathbf{r}$ the vector of rewards.

Consider the random variable

$$Y = \int_0^\tau r\,(X_t)\,dt.$$

Then, $Y$ is phase-type distributed (see [Kul89]).

Now consider $n$ nonegative vectors of rewards $\mathbf{r}_i = (r_i(1), \ldots, r_i(p))^\top$, $i = 1, \ldots, n$, which form the matrix

$$\mathbf{R} = \begin{pmatrix} \mathbf{r}_1 & \cdots & \mathbf{r}_n \end{pmatrix}.$$

Define

$$Y_i = \int_0^\tau |r_i\,(X_t)|\,dt$$

for every $i = 1, \ldots, n$.

Then, the random vector $(Y_1, \ldots, Y_n)$ is said to be multivariate phase-type distributed with parameters $(\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R})$ and we write $(Y_1, \ldots, Y_n) \sim \text{MPH}^*\,(\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R})$.

In Chapter 5, we present properties of the class of multivariate phase-type distributions defined above in the discrete time. One motivation for this work is that the literature of multivariate phase-type distributions has been mainly developed in the continuous time and actually there are a large number of multivariate phase-type distributions in the discrete time that may not be seen as part of this class of distributions.

## Overview of Concomitants of phase-type distributions.

In multivariate analysis, the concept of concomitant of order statistics comes from an ordering of bivariate random vectors. More precisely, if we consider the sample of bivariate random vectors

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_{n+1}, Y_{n+1}) \sim (X, Y)$$

and a random ordering of it with respect to one of the variates, let say from the $Y$-variates, then we have

$$\left(X_{[1:n+1]}, Y_{(1:n+1)}\right), \left(X_{[2:n+1]}, Y_{(2:n+1)}\right), \ldots, \left(X_{[n+1:n+1]}, Y_{(n+1:n+1)}\right),$$

where $Y_{(r:n+1)}$ denotes the $r$-th order statistics of the sample $Y_1, \ldots, Y_{n+1}$. Now, the paired variable with the $r$-th order statistics $Y_{(r:n+1)}$ is denoted by $X_{[r:n+1]}$ and is called the $r$-th concomitant or the concomitant of the $r$-th order statistic (see more in [Dav73]).

The theory of concomitants have been only developed for absolutely continuous bivariate random vector so far and mainly by considering a sample of independent and identically distributed bivariate random vectors. Few research has been done under the assumption of a sample of non-identically distributed but still independent random vectors (see [Ery05]) or dependent but identically distributed (see [WN09]). Concomitants have been widely applied in selection and prediction problems in regression analysis where is usually assumed bivariate normal distributions (see [Bha84], [DN98] and [HS79]). However, many more problems in multivariate analysis are related to concomitants and the simply inexistent knowledge on concomitants from multivariate phase-type distributions motivate us to look into this topic more deeply.

In the paper [TV17] a technique is established to uniquely determine a bivariate distribution by the density function of one of its marginals and the distribution of the concomitant of the largest or smaller order statistic. That analysis showed that the density function of concomitants can be written as a function of the density of one marginal and a semi-explicit Laplace transform of the bivariate distribution. This idea immediately reminded us the semi-explicit formula for a bivariate phase-type distribution given by L. Breuer in [Bre16].

In Chapter 6, we consider a sample of independent and identically phase-type distributed random vectors. We use the Kulkarni's representation for a bivariate phase-type distribution and the semi-explicit formula proposed by L. Breuer to show a procedure to calculate the distribution of concomitants and we prove that concomitants from bivariate phase-type distributions are phase-type distributed.

CHAPTER 2

# Matrix-geometric distributions

In this chapter, our aim is to provide the reader with the needed background on the discrete version of phase-type distributions and its discrete generalization called Matrix-geometric distributions, which helps to understand the content of the present thesis, specifically for Chapters 4 and 5. We start with the study of Discrete phase-type distributions and then we present the study of Matrix-geometric distributions, as it was historically introduced.

In the section of Discrete phase-type distributions, we choose to begin by presenting the underlying Markov chain and its properties, so right after we are able to present how a discrete phase-type distributed random variable is defined and give some typical examples of this distribution. Then, we talk about important properties of DPH-representations, these are its irreducibility and its non-uniqueness. We also present the formulas for the probability mass and distribution function together with the probability generating function and factorial moments of DPH-distributions. Lastly, this section ends with some closure properties of the class of DPH-distributions such as finite convolutions and finite mixtures.

For the section of Matrix-geometric distributions, we follow the same line than in the section of DPH-distributions, expect that here we start with the definition of a Matrix-geometric distribution. Then, we continue by presenting the formulas of the probability generating function, distribution function and factorial moments of this distribution and these formulas share to have the same form than the formulas in the DPH-distribution case. It has been shown a way to obtain MG-representations through the probability generating function of a MG-distribution, as well as the fact that we can always get a MG-representation with an exit vector which satisfies the same equation than an exit vector in the DPH-distribution case. As examples of MG-distributions, we believe that it is convenient for the following chapters to present the maximum and the minimum of two MG-distributed random variables. Finally, we present some closure properties of the class of MG-distributions, for instance finite convolutions, finite mixtures and N-fold convolutions.

## 2.1 Discrete Phase-type distributions

### 2.1.1 The underlying Markov chain and its properties

Consider a time-homogeneous Markov chain $\{X_n\}_{n\in\mathbb{N}}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a finite state space given by

$$\mathcal{E} = \{1, 2, \ldots, p, p+1\},$$

where $\{1, 2, \ldots, p\}$ are transient states and $p+1$ is absorbing. Let $\mathbf{P}$ denote the matrix of transition probabilities which is expressed on the form

$$\mathbf{P} = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}, \tag{2.1}$$

where $\mathbf{T}$ is the $p$-square matrix that contains the transition probabilities between the transient states, $\mathbf{t} = \mathbf{e} - \mathbf{T}\mathbf{e}$ is the vector of transition probabilities from the transient states to the absorbing state $p+1$ (sometimes referred to as the exit vector), and $\mathbf{e}$ denotes the $p$-dimensional column vector filled of ones.

The vector of initial probabilities of the Markov chain $\{X_n\}_{n\in\mathbb{N}}$ is denoted by $(\boldsymbol{\pi}, \pi_{p+1})$, where

$$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p),$$

and

$$\pi_i = \mathbb{P}(X_0 = i), i = 1, \ldots, p+1.$$

**Proposition 2.1** $\lim_{n\to\infty} \mathbf{T}^n = \mathbf{0}$.

**Proof.** Let $j$ be a transient state. Then, for every state $i \in \mathcal{E}$ it holds

$$\sum_{n=0}^{\infty} (\mathbf{P})_{ij}^n < \infty,$$

where

$$\mathbf{P}^n = \begin{pmatrix} \mathbf{T}^n & (\mathbf{I} - \mathbf{T}^n)\,\mathbf{e} \\ \mathbf{0} & 1 \end{pmatrix}$$

and $\mathbf{I}$ denotes the $p$-dimensional identity matrix (see Lemma A.5). Consequently

$$\lim_{n\to\infty} (\mathbf{P})_{ij}^n = 0.$$

In particular, due to the matrix $\mathbf{T}$ encloses all the possible transition probabilities between transient states, it follows that

$$\lim_{n\to\infty} \mathbf{T}^n = 0.$$

$\square$

Proposition 2.1 is equivalent to the next proposition.

Let $\rho\,(\mathbf{A})$ denotes the spectral radius of the matrix $\mathbf{A}$.

**Proposition 2.2** *Let* $\mathbf{T}$ *be the sub-transition probability matrix given in Equation* (2.1). *Then* $\rho\,(\mathbf{T}) < 1$.

A consequence of Proposition 2.1 is the next corollary.

**Corollary 2.3** *The matrix* $\mathbf{I} - \mathbf{T}$ *is nonsingular.*

See [Ste98, p. 55] for the proof.

### 2.1.2 Discrete phase-type distributions

Consider the stopping time

$$\tau = \min\{n \geq 0 : X_n = p + 1\},$$

this is the time until absorption of the Markov chain $\{X_n\}_{n\in\mathbb{N}}$.

**Definition 2.4** The distribution of $\tau$ is called **discrete phase-type distribution** (DPH-distribution) and the pair $(\boldsymbol{\pi}, \mathbf{T})$ is called a representation of the distribution of $\tau$. We write $\tau \sim \mathrm{DPH}_p\,(\boldsymbol{\pi}, \mathbf{T})$, where the number $p$ is said to be the dimension of the representation $(\boldsymbol{\pi}, \mathbf{T})$.

### 2.1.3 Examples

Three common examples of discrete phase-type distributions are given next to illustrate DPH-representations.

**Example 2.1 Geometric distribution.** *Let $W \sim$ Geometric$(\lambda)$. Then, $W \sim$* $\mathrm{DPH}_1\left(1, 1-\lambda\right)$.

**Example 2.2 Negative Binomial distribution.** *Let $Y \sim NB(m, 1-\lambda)$. Then, a discrete phase-type representation for this distribution is given by the $m$-dimensional row vector*

$$\boldsymbol{\pi} = (1, 0, 0, \ldots, 0, 0)\,,$$

*and the $m$-square matrix*

$$\mathbf{T} = \begin{pmatrix} 1-\lambda & \lambda & 0 & \cdots & 0 & 0 \\ 0 & 1-\lambda & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1-\lambda & \lambda \\ 0 & 0 & 0 & \cdots & 0 & 1-\lambda \end{pmatrix}.$$

*That is, $Y \sim \mathrm{DPH}_m\left(\boldsymbol{\pi}, \mathbf{T}\right)$.*

**Example 2.3 Generalized Negative Binomial distribution.** *Consider $m$ independent random variables given by $W_i \sim$ Geometric$(\lambda_i)$, $i = 1, \ldots, m$. Let $Y = \sum_{i=1}^{m} W_i$. Then, $Y$ is discrete phase-type distributed with a representation given by an $m$-dimensional row vector*

$$\boldsymbol{\pi} = (1, 0, 0, \ldots, 0, 0)\,,$$

*and an $m$-square matrix*

$$\mathbf{T} = \begin{pmatrix} 1-\lambda_1 & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & 1-\lambda_2 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1-\lambda_{m-1} & \lambda_{m-1} \\ 0 & 0 & 0 & \cdots & 0 & 1-\lambda_m \end{pmatrix}.$$

### 2.1.4 Properties of DPH-representations

In the case where $\pi_{p+1} > 0$, it says that the distribution of $\tau$ has a **zero point of size** $\pi_{p+1}$.

Throughout this chapter, we usually assume that $\pi_{p+1} = 0$, unless is specified the other case.

Notice from Corollary 2.3 that

$$\begin{aligned}
(\mathbf{I} - \mathbf{T})^{-1}_{i,j} &= \sum_{n=0}^{\infty} (\mathbf{T})^n_{ij} \\
&= \sum_{n=0}^{\infty} \mathbb{P}\left( X_n = j, \tau > n \middle| X_0 = i \right) \\
&= \mathbb{E}\left( \sum_{n=0}^{\tau-1} \mathbf{1}_{\{X_n = j\}} \middle| X_0 = i \right).
\end{aligned}$$

Thus, $(\mathbf{I} - \mathbf{T})^{-1}_{i,j}$ represents the expected number of times the underlying Markov chain $\{X_n\}_{n\in\mathbb{N}}$ visits the state $j$ before it gets absorbed and starting from state $i$.

**Definition 2.5** A DPH-representation $(\boldsymbol{\pi}, \mathbf{T})$ is said to be **irreducible** if and only if the matrix $\mathbf{T} + \mathbf{t}\boldsymbol{\pi}$ is an irreducible matrix.

For every discrete phase-type distribution we can construct an irreducible DPH-representation from the original one (see [Neu75, p. 189]). Therefore, we also assume that all the DPH-representations present in this thesis are irreducible, unless we specify the other case.

**Non-uniqueness of the representation.**
In Example 2.3, observe that by considering any other permutation of the variables $W_i, i = 1, \ldots, m$ in the sum $Y$, it generates another representation for the distribution. Therefore, in general, the representation for a discrete phase-type distribution is not unique. Furthermore, in the following is shown that representations for the same distribution can have different dimension.

**Example 2.4** *Consider* $\tau \sim \mathrm{DPH}_p(\boldsymbol{\pi}, \mathbf{T})$, *where* $\mathbf{T}$ *is irreducible. Since* $\mathbf{T}$ *is a sub-stochastic matrix, then the spectrum radius of* $\mathbf{T}$ *is lower than one. By the Perron–Frobenius theorem,* $\mathbf{T}$ *has a left eigenvector* $\mathbf{v}$, *with eigenvalue* $1 - \lambda > 0$ *such as* $\lambda < 1$. *Let* $\mathbf{v}^*$ *be the vector obtained by normalizing the eigenvector* $\mathbf{v}$, *this is* $\mathbf{v}^*\mathbf{e} = 1$. *Then*

$$\begin{aligned}
\mathbf{v}^*\mathbf{T}^m\mathbf{t} &= (1-\lambda)\mathbf{v}^*\mathbf{T}^{m-1}\mathbf{t} \\
&= (1-\lambda)^m \mathbf{v}^*\mathbf{t} \\
&= (1-\lambda)^m (1 - \mathbf{v}^*\mathbf{T}\mathbf{e}) \\
&= (1-\lambda)^m \lambda, \quad m \in \mathbb{N}.
\end{aligned}$$

*Therefore, the representation* $(\mathbf{v}^*, \mathbf{T})$ *of dimension* $p$ *is another discrete phase-type representation for the geometric distribution with parameter* $\lambda$.

From Example 2.3 we can notice that even though the DPH-representation is irreducible, that can be not unique.

**Definition 2.6** Let $\tau$ be a discrete phase-type distributed random variable and let $\mathcal{R}_\tau$ denote the set of representations for the distribution of $\tau$. The smallest dimension of the representations in $\mathcal{R}_\tau$ is called **the order of the distribution of** $\tau$.

### 2.1.5 Properties of DPH-distributions

In this part we present the formulas for the probability mass and distribution function of DPH-distributions, as well as its probability generating function and factorial moments.

**Proposition 2.7** *Let* $\tau \sim \mathrm{DPH}_p(\boldsymbol{\pi}, \mathbf{T})$ *. Then the following statements hold.*

*i. The probability distribution of* $\tau$, $F_\tau$, *is given by*

$$F_\tau(m) = 1 - \boldsymbol{\pi}\mathbf{T}^m\mathbf{e}, \quad m \in \mathbb{N}_0. \tag{2.2}$$

*ii. The probability mass function of* $\tau$, $f_\tau$, *is given by*

$$f_\tau(m) = \boldsymbol{\pi}\mathbf{T}^{m-1}\mathbf{t}, \quad m \in \mathbb{N}_0. \tag{2.3}$$

*iii. The probability generating function (p.g.f.) of* $\tau$ *is*

$$\mathbb{E}(s^\tau) = s\boldsymbol{\pi}(\mathbf{I} - s\mathbf{T})^{-1}\mathbf{t}, \tag{2.4}$$

*for* $|s| \leq 1$.

*iv. The expectation of* $\tau$ *is*

$$\mathbb{E}(\tau) = \boldsymbol{\pi}(\mathbf{I} - \mathbf{T})^{-1}\mathbf{e}. \tag{2.5}$$

*v. For every* $n \in \mathbb{N}$, *the factorial moment of* $\tau$ *is given by*

$$\mathbb{E}\big(\tau(\tau-1)\cdots(\tau-(n-1))\big) = n!\boldsymbol{\pi}\mathbf{T}^{n-1}(\mathbf{I} - \mathbf{T})^{-n}\mathbf{e}. \tag{2.6}$$

**Proof.** $i$. Observe that

$$1 - F_\tau(m) = \sum_{j=1}^{p} \mathbb{P}(X_m = j) = \sum_{i,j=1}^{p} \mathbb{P}(X_m = j | X_0 = i)\,\mathbb{P}(X_0 = i)$$

$$= \quad \boldsymbol{\pi}\mathbf{T}^m\mathbf{e}.$$

Therefore $F_\tau(m) = 1 - \boldsymbol{\pi}\mathbf{T}^m\mathbf{e}$.

$ii.$

$$\begin{aligned}
f_\tau(m) &= \sum_{j=1}^{p} \mathbb{P}\left(X_m = p+1, X_{m-1} = j\right) \\
&= \sum_{j=1}^{p} \mathbb{P}\left(X_m = p+1 | X_{m-1} = j\right) \mathbb{P}\left(X_{m-1} = j\right) \\
&= \sum_{i,j=1}^{p} \mathbb{P}\left(X_m = p+1 | X_{m-1} = j\right) \mathbb{P}\left(X_{m-1} = j | X_0 = i\right) \mathbb{P}\left(X_0 = i\right) \\
&= \boldsymbol{\pi}\mathbf{T}^{m-1}\mathbf{t}.
\end{aligned}$$

$iii.$

$$\begin{aligned}
\mathbb{E}\left(s^\tau\right) &= \sum_{m=1}^{\infty} s^m \boldsymbol{\pi}\mathbf{T}^{m-1}\mathbf{t} \\
&= s\boldsymbol{\pi}\left(\sum_{m=0}^{\infty} \left(s\mathbf{T}\right)^m\right)\mathbf{t} \\
&= s\boldsymbol{\pi}\left(\mathbf{I} - s\mathbf{T}\right)^{-1}\mathbf{t} \quad \text{for} \quad |s| \leq 1.
\end{aligned}$$

$iv.$

$$\begin{aligned}
\mathbb{E}\left(\tau\right) &= \sum_{m=0}^{\infty} \mathbb{P}\left(\tau > m\right) \\
&= \sum_{m=0}^{\infty}\sum_{i=1}^{p} \mathbb{P}\left(\tau > m | X_0 = i\right) \mathbb{P}\left(X_0 = i\right) \\
&= \sum_{m=0}^{\infty}\sum_{i=1}^{p} \pi_i \left(\mathbf{T}^m\mathbf{e}\right)_i \\
&= \sum_{i=1}^{p} \pi_i \sum_{m=0}^{\infty} \left(\mathbf{T}^m\mathbf{e}\right)_i \\
&= \sum_{i=1}^{p} \pi_i \left[\left(\sum_{m=0}^{\infty} \mathbf{T}^m\right)\mathbf{e}\right]_i \\
&= \sum_{i=1}^{p} \pi_i \left[\left(\mathbf{I} - \mathbf{T}\right)^{-1}\mathbf{e}\right]_i
\end{aligned}$$

$$= \boldsymbol{\pi} \left(\mathbf{I} - \mathbf{T}\right)^{-1} \mathbf{e}.$$

$v$. We recall that

$$\mathbb{E}\left(\prod_{i=1}^{n} (\tau - i + 1)\right) = \left.\frac{d^n \mathbb{E}(s^\tau)}{ds^n}\right|_{s=1}.$$

Then the proof is completed by induction on the $n$-derivative of the probability-generating function of $\tau$. Specifically, by proving that

$$\frac{d^n}{ds^n} \mathbb{E}(s^\tau) = n! \boldsymbol{\pi} \mathbf{T}^{n-1} \left(\mathbf{I} - s\mathbf{T}\right)^{-n-1} \mathbf{t},$$

for every $n \in \mathbb{N}$. $\square$

### 2.1.6 Closure properties of the class of DPH-distributions

**Theorem 2.8 (Finite convolution of DPH-distributions.)** *The convolution of a finite number of probability mass functions of discrete phase-type is itself a probability mass function of discrete phase-type.*

The proof is originally given in [Neu75, p. 177] and it is exact the same than the proof of Property 2.1.

**Theorem 2.9 (Finite Mixtures of DPH-distributions.)** *Any finite mixture of probability mass functions of discrete phase-type is itself of a probability mass functions of discrete phase-type.*

The proof is the same than the proof in Property 2.2 (see also [Neu75, p. 179]).

In general, an infinite mixture of probability discrete phase-type distributions is not discrete phase-type distributed. However, M. F. Neuts in [Neu75, p. 181] provided an class of infinite mixtures of discrete phase-type distributions that are discrete phase-type. The result is given next.

**Theorem 2.10 (The N-fold convolution.)** *Let* $N \sim \mathrm{DPH}_q\left(\boldsymbol{\beta}, \mathbf{B}\right)$ *and let* $\{W_i\}_{i \in \mathbb{N}}$ *be a sequence of independent and identically discrete phase-type distributed random variables with representation* $\mathrm{DPH}_p\left(\boldsymbol{\pi}, \mathbf{T}\right)$*. Then, the compound sum*

$$\sum_{i=1}^{N} W_i$$

*is discrete phase-type distributed with a representation given by*

$$(\boldsymbol{\pi} \otimes \boldsymbol{\beta}, \mathbf{T} \otimes \mathbf{I}_q + \mathbf{t}\boldsymbol{\pi} \otimes \mathbf{B}),$$

*where* $\mathbf{I}_q$ *is the identity matrix of dimension* $q$ *and* $\mathbf{t} = \mathbf{e} - \mathbf{T}\mathbf{e}$.

**Theorem 2.11** *[Order statistics from DPH-distributions.] Order statistics from a finite set of independent and DPH-distributed random variables are DPH-distributed.*

In Section 4.4, we present the proof of Theorem 2.11.

**Theorem 2.12** *Any probability mass function* $f$ *with finite support on* $\mathbb{N}$ *is a discrete phase-type.*

This result is given in [Neu75, p. 177]. The proof is simple since we can construct an absorbing Markov chain by matching every probability of a point of $f$ with the probability of the Markov chain being absorbed at the corresponding point of time. An example of this construction is presented in [Gre09, p. 14].

**Theorem 2.13 (Characterization of Discrete Phase-type Distribution.)**
*A probability measure* $\mu$ *on* $\{0, 1, \ldots\}$ *with rational generating function is of phase-type if and only if for some positive integer* $\mathrm{w}$ *each of the generating functions*

$$f_i(z) = \sum_{k=0}^{\infty} \mu(\mathrm{w}k + i)\, z^k,$$

*where for* $i = 0, 1, \ldots, \mathrm{w} - 1$, *is either a polynomial or has a unique pole of minimal absolute value.*

In order to recall the proof, we refer to [O'C90b].

## 2.2 Matrix-geometric distributions

### 2.2.1 Definition

Consider an $p$-dimensional row vector $\boldsymbol{\varsigma}$, an $p$-square matrix $\mathbf{Z}$ and an $p$-dimensional column vector $\mathbf{z}$ taking values on $\mathbb{C}$ and such that the function

$$g(m) = \boldsymbol{\varsigma}\mathbf{Z}^{m-1}\mathbf{z}, m \in \mathbb{N}, \quad g(0) = g_0, \tag{2.7}$$

is a probability mass function.

A random variable $X$ having a probability mass function of the form (2.7) is said to be **Matrix-Geometric distributed** (MG-distributed) with representation $(\boldsymbol{\varsigma}, \mathbf{Z}, \mathbf{z})$ . We write $X \sim \mathrm{MG}_p(\boldsymbol{\varsigma}, \mathbf{Z}, \mathbf{z})$ , where the number $p$ is the dimension of the representation (also called the order of the representation). The vector $\mathbf{z}$ is referred to as "the exit vector", which is adopted from the case of DPH-distributions.

Here, in many cases, we do not mention the dimension of the representation when is not required for the main purpose.

If $g_0 \neq 0$, the distribution of $X$ has a zero point of size $g_0$. To simplify formulations, we assume that all matrix-geometric distributed random variables presented here do not have a zero point, unless it is specified.

### 2.2.2 Properties of MG-distributions.

Let $X \sim \mathrm{MG}(\boldsymbol{\varsigma}, \mathbf{Z}, \mathbf{z})$ . In order to calculate the probability-generating function of $X$, observe that

$$\sum_{n=0}^{\infty} (s\mathbf{Z})^n$$

coverges if and only if $|s| < \frac{1}{\rho(\mathbf{Z})}$, where $\rho(\mathbf{Z})$ is the spectral radius of the matrix $\mathbf{Z}$. Therefore, if $|s| < \frac{1}{\rho(\mathbf{Z})}$, then the probability generating function is well-defined and is given by

$$\mathbb{E}\left(s^X\right) = s\boldsymbol{\varsigma}\left(\mathbf{I} - s\mathbf{Z}\right)^{-1}\mathbf{z}. \tag{2.8}$$

**Proposition 2.14** *Let $X \sim \mathrm{MG}(\boldsymbol{\varsigma}, \mathbf{Z}, \mathbf{z})$ and assume that $\rho(\mathbf{Z}) < 1$. Then*

*i. The cumulative distribution function of $X$ is given by*

$$\mathbb{P}(X \leq m) = 1 - \boldsymbol{\varsigma}\mathbf{Z}^m\left(\mathbf{I} - \mathbf{Z}\right)^{-1}\mathbf{z}. \tag{2.9}$$

*ii. The expectation of $X$ is given by*

$$\mathbb{E}(X) = \boldsymbol{\varsigma}\left(\mathbf{I} - \mathbf{Z}\right)^{-2}\mathbf{z}. \tag{2.10}$$

*iii. The factorial moments of $X$ are given by*

$$\mathbb{E}\left(\prod_{i=1}^{n}(X - i + 1)\right) = n!\boldsymbol{\varsigma}\mathbf{Z}^{n-1}\left(\mathbf{I} - \mathbf{Z}\right)^{-n-1}\mathbf{z}. \tag{2.11}$$

**Proof.** $i.$

$$\begin{aligned}\mathbb{P}(X \leq m) &= \sum_{i=1}^{m} \mathbb{P}(X = i) \\ &= \varsigma \left( \sum_{i=1}^{m} \mathbf{Z}^{i-1} \right) \mathbf{z} \\ &= \varsigma \left( (\mathbf{I} - \mathbf{Z}^{m}) (\mathbf{I} - \mathbf{Z})^{-1} \right) \mathbf{z} \\ &= 1 - \varsigma \mathbf{Z}^{m} (\mathbf{I} - \mathbf{Z})^{-1} \mathbf{z}.\end{aligned}$$

$ii.$ The expectation of $X$ can be derived as follows

$$\begin{aligned}\mathbb{E}(X) &= \sum_{m=1}^{\infty} \mathbb{P}(X \geq m) \\ &= \sum_{m=1}^{\infty} (1 - \mathbb{P}(X \leq m-1)) \\ &= \varsigma \sum_{m=1}^{\infty} \mathbf{Z}^{m-1} (\mathbf{I} - \mathbf{Z})^{-1} \mathbf{z} \\ &= \varsigma (\mathbf{I} - \mathbf{Z})^{-1} (\mathbf{I} - \mathbf{Z})^{-1} \mathbf{z} \quad \text{by Equation (2.9)} \\ &= \varsigma (\mathbf{I} - \mathbf{Z})^{-2} \mathbf{z},\end{aligned}$$

$iii.$ The proof is exactly the same than in the case of DPH-distributions. That is by calculating

$$\mathbb{E}\left( \prod_{i=1}^{n} (X - i + 1) \right) = \left. \frac{d^n \mathbb{E}\left(s^X\right)}{ds^n} \right|_{s=1}.$$

□

### 2.2.3 Properties of MG-representations

The class of matrix-geometric distributions has a rational probability generating function from which is derived a practical way to obtain a MG-representation. The result is in the following Proposition.

**Proposition 2.15** *Let $X$ be a matrix-geometric distributed random variable and assume that its probability generating function is given by*

$$\mathbb{E}\left(s^X\right) = \frac{b_n s + b_{n-1} s^2 + \cdots + b_3 s^{n-2} + b_2 s^{n-1} + b_1 s^n}{1 + a_1 s + a_2 s^2 + \cdots + a_{n-2} s^{n-2} + a_{n-1} s^{n-1} + a_n s^n} \tag{2.12}$$

*for some $n \in \mathbb{N}$ and constants $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$. Then, the density function of $X$ is calculated as*

$$\mathbb{P}(X = m) = \boldsymbol{\gamma}\mathbf{G}^{m-1}\mathbf{g}, \quad m \in \mathbb{N},$$

*where $\boldsymbol{\gamma}$ is an $n$-dimensional row vector given by*

$$\begin{aligned} \boldsymbol{\gamma} &= (1, 0, \ldots, 0), \\ \mathbf{G} &= \begin{pmatrix} -a_1 & 0 & 0 & \cdots & 0 & 1 \\ -a_n & 0 & 0 & \cdots & 0 & 0 \\ -a_{n-1} & 1 & 0 & \cdots & 0 & 0 \\ -a_{n-2} & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -a_2 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \end{aligned} \tag{2.13}$$

*and*

$$\mathbf{g} = \begin{pmatrix} b_n \\ b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{n-2} \\ b_{n-1} \end{pmatrix}.$$

For the proof we refer [Gre09, Proposition 4.2].

**Proposition 2.16** *Let $(\boldsymbol{\varsigma}, \mathbf{Z}, \mathbf{z})$, where $\rho(\mathbf{Z}) < 1$, be a matrix-geometric representation for the distribution of the random variable $X$. Then, based on the representation $(\boldsymbol{\varsigma}, \mathbf{Z}, \mathbf{z})$, there exists another matrix-geometric representation for the distribution of $X$, $(\boldsymbol{\varrho}, \mathbf{Y}, \mathbf{y})$, such that $\mathbf{y} = \mathbf{e} - \mathbf{Y}\mathbf{e}$.*

**Proof.** Let $\mathbf{M}$ be a nonsingular matrix. Then the density function of $X$ can be written as

$$\begin{aligned} \mathbb{P}(X = m) &= \boldsymbol{\varsigma}\mathbf{Z}^{m-1}\mathbf{z} \\ &= \boldsymbol{\varsigma}\mathbf{M}\left(\mathbf{M}^{-1}\mathbf{Z}^{m-1}\mathbf{M}\right)\mathbf{M}^{-1}\mathbf{z} \\ &= \boldsymbol{\varsigma}\mathbf{M}\left(\mathbf{M}^{-1}\mathbf{Z}\mathbf{M}\right)^{m-1}\mathbf{M}^{-1}\mathbf{z}. \end{aligned}$$

If the matrix $\mathbf{Y}$ is chosen to be equal to $\mathbf{M}^{-1}\mathbf{Z}\mathbf{M}$, then $\mathbf{M}$ should satisfy the equation

$$\mathbf{M}^{-1}\mathbf{Z}\mathbf{M}\mathbf{e} + \mathbf{M}^{-1}\mathbf{z} = \mathbf{e},$$

which is equivalent to

$$\mathbf{M}\mathbf{e} = (\mathbf{I} - \mathbf{Z})^{-1}\mathbf{z}. \tag{2.14}$$

In the next, the nonsingular matrix $\mathbf{M}$ is constructed based on the condition (2.14). For that aim, let us assume that $(\varsigma, \mathbf{Z}, \mathbf{z})$ is the canonical form given in (2.13) of the distribution of $X$. First, from the probability-generating function, we have

$$\mathbb{E}\left(1^X\right) = \varsigma\left(\mathbf{I}-\mathbf{Z}\right)^{-1}\mathbf{z} = 1,$$

which implies that the first entry of the vector $(\mathbf{I}-\mathbf{Z})^{-1}\mathbf{z}$ is equal to one (see Equation (2.13)). Now, we can form the matrix $\mathbf{M}$ as follows.

For the first row of $\mathbf{M}$ we have

$$\begin{aligned} \mathbf{M}_{1,1} &= 1, \\ \mathbf{M}_{1,j} &= 0 \quad \text{for all} \quad 2 \le j \le n. \end{aligned}$$

For the $i$-th row of $\mathbf{M}$, where $2 \le i \le n-1$, it depends on the value of the $i$-th entry of the vector $(\mathbf{I}-\mathbf{Z})^{-1}\mathbf{z}$.

If $\left((\mathbf{I}-\mathbf{Z})^{-1}\mathbf{z}\right)_i \neq 0$, then the $i$-th row of the matrix $\mathbf{M}$ is defined by

$$\begin{aligned} \mathbf{M}_{i,i} &= \left((\mathbf{I}-\mathbf{Z})^{-1}\mathbf{z}\right)_i, \\ \mathbf{M}_{i,j} &= 0 \quad \text{for all} \quad j \neq i. \end{aligned}$$

If $\left((\mathbf{I}-\mathbf{Z})^{-1}\mathbf{z}\right)_i = 0$, then the $i$-th row of the matrix $\mathbf{M}$ is defined

$$\begin{aligned} \mathbf{M}_{i,i} &= 1, \\ \mathbf{M}_{i,i+1} &= -1, \\ \mathbf{M}_{i,j} &= 0 \quad \text{other case.} \end{aligned}$$

For the $n$-th row of $\mathbf{M}$.

If $\left((\mathbf{I}-\mathbf{Z})^{-1}\mathbf{z}\right)_n \neq 0$, then the $n$-th row of the matrix $\mathbf{M}$ is defined by

$$\begin{aligned} \mathbf{M}_{n,n} &= \left((\mathbf{I}-\mathbf{Z})^{-1}\mathbf{z}\right)_n, \\ \mathbf{M}_{i,j} &= 0 \quad \text{for all} \quad j \neq i. \end{aligned}$$

If $\left((\mathbf{I}-\mathbf{Z})^{-1}\mathbf{z}\right)_n = 0$, then the $n$-th row of the matrix $\mathbf{M}$ is defined

$$\mathbf{M}_{n,n} = 1,$$

$$\begin{aligned}\mathbf{M}_{n,1} &= -1,\\ \mathbf{M}_{i,j} &= 0 \quad \text{other case.}\end{aligned}$$

In this way the $n$ rows are linearly independent, which implies that $\mathbf{M}$ is nonsingular.

□

Since we can always assume that from one MG-representation we can find another one $(\boldsymbol{\varrho}, \mathbf{Y}, \mathbf{y})$ such that $\mathbf{y} = \mathbf{e} - \mathbf{Y}\mathbf{e}$, then the cumulative distribution function, Expectation and factorial moments given in Equations (2.9), (2.10) and (2.11), respectively, are reduced to the form

$$\begin{aligned}\mathbb{P}\left(X \leq m\right) &= 1 - \boldsymbol{\varrho}\mathbf{Y}^{m}\mathbf{e},\\ \mathbb{E}\left(X\right) &= \boldsymbol{\varrho}\left(\mathbf{I} - \mathbf{Y}\right)^{-1}\mathbf{e},\\ \mathbb{E}\left(\prod_{i=1}^{n}\left(X - i + 1\right)\right) &= n!\boldsymbol{\varrho}\mathbf{Y}^{n-1}\left(\mathbf{I} - \mathbf{Y}\right)^{-n}\mathbf{e},\end{aligned}$$

which coincide with the corresponding formulas in the case of DPH-distributions.

### 2.2.4 Examples

For the general idea of this thesis, it is important to show particular cases of order statistics from matrix-geometric distributions. For that reason, in the following we introduce the maximum and the minimum of two independent and matrix-geometric distributed random variables as examples of matrix-geometric distributions.

**Example 2.5 (Maximum and Minimum.)** *Let $Y_1$ and $Y_2$ be two independent and MG-distributed random variables with representations given by $(\boldsymbol{\varsigma}_1, \mathbf{Z}_1, \mathbf{z}_1)$ and $(\boldsymbol{\varsigma}_2, \mathbf{Z}_2, \mathbf{z}_2)$, respectively.*

**The Maximum of $Y_1$ and $Y_2$.**

*We notice that the survival function of the distribution of the maximum of $Y_1$ and $Y_2$ can be calculated as follows.*

$$\begin{aligned}\mathbb{P}\left(\max(Y_1, Y_2) > m\right) &= 1 - \mathbb{P}\left(Y_1 \leq m\right)\mathbb{P}\left(Y_2 \leq m\right)\\ &= \boldsymbol{\varsigma}_1\mathbf{Z}_1^{m}\mathbf{e} + \boldsymbol{\varsigma}_2\mathbf{Z}_2^{m}\mathbf{e} - \left(\boldsymbol{\varsigma}_1 \otimes \boldsymbol{\varsigma}_2\right)\left(\mathbf{Z}_1 \otimes \mathbf{Z}_2\right)^{m}\mathbf{e}\\ &= \left(-\left(\boldsymbol{\varsigma}_1 \otimes \boldsymbol{\varsigma}_2\right), \boldsymbol{\varsigma}_1, \boldsymbol{\varsigma}_2\right)\begin{pmatrix}\mathbf{Z}_1 \otimes \mathbf{Z}_2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_2\end{pmatrix}^{m}\mathbf{e}.\end{aligned}$$

*Then, a MG-representation for the distribution of* $\max(Y_1, Y_2)$ *is given by* $\left(\boldsymbol{\varsigma}_{2:2}, \mathbf{Z}_{2:2}, \mathbf{z}_{2:2}\right)$, *where*

$$\boldsymbol{\varsigma}_{2:2} = \left(-\left(\boldsymbol{\varsigma}_1 \otimes \boldsymbol{\varsigma}_2\right), \boldsymbol{\varsigma}_1, \boldsymbol{\varsigma}_2\right), \quad \mathbf{Z}_{2:2} = \begin{pmatrix} \mathbf{Z}_1 \otimes \mathbf{Z}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_2 \end{pmatrix}$$

*and* $\mathbf{z}_{2:2} = \mathbf{e} - \mathbf{Z}_{2:2}\mathbf{e}$.

**The Minimum of $Y_1$ and $Y_2$.**

*Assume that* $\mathbf{z}_i = \mathbf{e} - \mathbf{Z}_i, i = 1, 2$. *Then, the survival function of the distribution of the minimum of* $Y_1$ *and* $Y_2$ *is given by*

$$\begin{aligned} \mathbb{P}\left(\min\left(Y_1, Y_2\right) > m\right) &= \mathbb{P}\left(Y_1 > m\right) \mathbb{P}\left(Y_2 > m\right) \\ &= \left(\boldsymbol{\varsigma}_1 \otimes \boldsymbol{\varsigma}_2\right) \left(\mathbf{Z}_1 \otimes \mathbf{Z}_2\right)^m \left(\mathbf{e} \otimes \mathbf{e}\right). \end{aligned}$$

*Therefore, a MG-representation for the distribution of* $\min(Y_1, Y_2)$ *is*

$$\left(\left(\boldsymbol{\varsigma}_1 \otimes \boldsymbol{\varsigma}_2\right), \left(\mathbf{Z}_1 \otimes \mathbf{Z}_2\right)\right),$$

*and the exit vector is given by* $\left(\mathbf{e} \otimes \mathbf{e}\right) - \left(\mathbf{Z}_1 \otimes \mathbf{Z}_2\right)\left(\mathbf{e} \otimes \mathbf{e}\right)$.

### 2.2.5 Properties of the class of MG-distributions

**Theorem 2.17** *The class of matrix-geometric distributions is strictly larger than the class of discrete phase-type distributions.*

The proof consists on an example of a genuine MG-distribution, this is a distribution with a rational probability generating function that does not belong to the class of discrete phase-type distributions according to the characterization given in Theorem 2.13. For the proof we refer [Gre09, p. 36].

#### 2.2.5.1 Closure properties of the class of MG-distributions

**Property 2.1 (Finite convolutions of MG-distributions.)** *The convolution of a finite number of MG-distributions is itself MG-distributed.*

**Proof.** Let $X$ and $Y$ be two independent and MG-distributed random variables with representation given by $\mathrm{MG}_p\left(\boldsymbol{\varsigma}_1, \mathbf{Z}_1, \mathbf{z}_1\right)$ and $\mathrm{MG}_q\left(\boldsymbol{\varsigma}_2, \mathbf{Z}_2, \mathbf{z}_2\right)$, respectively. Consider the row vector

$$\boldsymbol{\varphi} = \left(\boldsymbol{\varsigma}_1, \mathbf{0}_q\right),$$

where $\mathbf{0}_q$ is a row vector filled of zeros of dimension $q$, consider also the matrix

$$\mathbf{U}=\begin{pmatrix}\mathbf{Z}_1 & \mathbf{z}_1\boldsymbol{\varsigma}_2\\ \mathbf{0}_{q\times p} & \mathbf{Z}_2\end{pmatrix}$$

where $\mathbf{0}_{q\times p}$ is a matrix with zero-valued entries of dimension $q\times p$, and, lastly, consider the column vector

$$\mathbf{u}=\begin{pmatrix}\mathbf{0}_p\\ \mathbf{z}_2\end{pmatrix},$$

where $\mathbf{0}_p$ is a column vector filled of zeros of dimension $p$. We recall formula (2.8). Then, for $|s|<\frac{1}{\rho(\mathbf{U})}$ we calculate the following

$$\begin{aligned}
s\boldsymbol{\varphi}\left(\mathbf{I}-s\mathbf{U}\right)^{-1}\mathbf{u} &= s\boldsymbol{\varphi}\begin{pmatrix}\mathbf{I}_p-s\mathbf{Z}_1 & -s\mathbf{z}_1\boldsymbol{\varsigma}_2\\ \mathbf{0}_{q\times p} & \mathbf{I}_q-s\mathbf{Z}_2\end{pmatrix}^{-1}\mathbf{u}\\
&= s\boldsymbol{\varphi}\begin{pmatrix}\left(\mathbf{I}_p-s\mathbf{Z}_1\right)^{-1} & s\left(\mathbf{I}_p-s\mathbf{Z}_1\right)^{-1}\mathbf{z}_1\boldsymbol{\varsigma}_2\left(\mathbf{I}_q-s\mathbf{Z}_2\right)^{-1}\\ \mathbf{0}_{q\times p} & \left(\mathbf{I}_q-s\mathbf{Z}_2\right)^{-1}\end{pmatrix}\mathbf{u}\\
&= \left(s\boldsymbol{\varsigma}_1\left(\mathbf{I}_p-s\mathbf{Z}_1\right)^{-1}, s^2\boldsymbol{\varsigma}_1\left(\mathbf{I}_p-s\mathbf{Z}_1\right)^{-1}\mathbf{z}_1\boldsymbol{\varsigma}_2\left(\mathbf{I}_q-s\mathbf{Z}_2\right)^{-1}\right)\mathbf{u}\\
&= \left(s\boldsymbol{\varsigma}_1\left(\mathbf{I}_p-s\mathbf{Z}_1\right)^{-1}\mathbf{z}_1\right)\left(s\boldsymbol{\varsigma}_2\left(\mathbf{I}_q-s\mathbf{Z}_2\right)^{-1}\mathbf{z}_2\right)\\
&= \mathbb{E}\left(e^{-sX}\right)\mathbb{E}\left(e^{-sY}\right)=\mathbb{E}\left(e^{-s(X+Y)}\right),
\end{aligned}$$

where the second equality follows by Lemma A.4. Therefore, $(\boldsymbol{\varphi},\mathbf{U},\mathbf{u})$ is a MG-representation for $X+Y$. □

**Property 2.2 (Finite Mixtures of MG-distributions.)** *Any finite mixture of MG-distributions is itself MG-distributed.*

**Proof.** Let $X$ and $Y$ be two MG-distributed random variables with representation given by $\mathrm{MG}_p\left(\boldsymbol{\varsigma}_1,\mathbf{Z}_1,\mathbf{z}_1\right)$ and $\mathrm{MG}_q\left(\boldsymbol{\varsigma}_2,\mathbf{Z}_2,\mathbf{z}_2\right)$, respectively, and let $\theta\in[0,1]$. Denote the corresponding densities as follows

$$f_X(m)=\boldsymbol{\varsigma}_1\mathbf{Z}_1^{m-1}\mathbf{z}_1 \text{ and } f_Y(m)=\boldsymbol{\varsigma}_2\mathbf{Z}_2^{m-1}\mathbf{z}_2,\quad m\in\mathbb{N}.$$

Then the mixture given by

$$\theta f_X(m)+(1-\theta)f_Y(m)$$

has a MG-representation $(\boldsymbol{\varphi},\mathbf{U},\mathbf{u})$ where

$$\boldsymbol{\varphi}=\left(\theta\boldsymbol{\varsigma}_1,(1-\theta)\boldsymbol{\varsigma}_2\right),\quad \mathbf{U}=\begin{pmatrix}\mathbf{Z}_1 & \mathbf{0}_{p\times q}\\ \mathbf{0}_{q\times p} & \mathbf{Z}_2\end{pmatrix} \text{ and } \mathbf{u}=\begin{pmatrix}\mathbf{z}_1\\ \mathbf{z}_2\end{pmatrix}.$$

□

**Property 2.3 (Order statistics from MG-distributions.)** *Order statistics from independent and MG-distributed random variables are MG-distributed.*

We present the simplest cases in Example 2.5 and the proof for the general case is given in Section 4.5.

**Property 2.4 (N-fold convolution.)** *Let* $\{Y_n\}_{n\in\mathbb{N}}$ , *where* $Y_n \sim \mathrm{MG}(\boldsymbol{\pi}, \mathbf{T}, \mathbf{t})$ , *be a sequence of independent and identically MG-distributed random variables and let* $N \sim \mathrm{MG}(\boldsymbol{\varsigma}, \mathbf{Z}, \mathbf{z})$ . *Consider the compound sum given by*

$$W = \sum_{n=1}^{N} Y_n.$$

*Then,* $W$ *is MG-distributed with a representation* $(\boldsymbol{\varrho}, \mathbf{Y}, \mathbf{y})$ , *where*

$$\begin{aligned}\boldsymbol{\varrho} &= \boldsymbol{\pi} \otimes \mathbf{z},\\ \mathbf{Y} &= \mathbf{T} \otimes \mathbf{I} + (\mathbf{t}\mathbf{z}) \otimes \mathbf{Z},\\ \mathbf{y} &= \mathbf{t} \otimes \boldsymbol{\zeta}.\end{aligned}$$

The same analytic proof provided for the case of DPH-distributions in [Neu75, p. 182] can be applied for the case of MG-distributions.

CHAPTER 3

# Matrix-Exponential distributions

This chapter provides a basic introduction on Phase-type and Matrix-exponential distributions to facilitate the understanding of Chapters 4 and 6. We start by introducing first phase-type distributions and then Matrix-exponential distributions, as it was historically introduced.

On the Phase-type distribution part, we talk about the particular structure of the underlying Markov jump process and after have established that base, we introduce the definition of phase-type distribution as the distribution of a stopping time of the underlying Markov jump process. We discuss irreducible PH-representations and the most important properties of the distribution such as the closed form formula of its density and cumulative distribution function, as well as the Laplace transform and moments of the distribution. Also, we present some of the most common and easy examples of Phase-type distributions, for instance the Exponential and Erlang distribution. Lastly, we finish this part with important properties of the class of Phase-type distributions, these are the characterization theorem for phase-type distributions and some closure properties such as finite convolutions and mixtures, which are useful to construct representations of phase-type in many applications.

In the section of Matrix-exponential distributions, we have followed the same line of analysis than in the part of phase-type distributions. We start with the definition of Matrix-exponential distributions to posteriorly state its properties such as the cumulative distribution function, the Laplace transform and the moments. Then, we present some properties of ME-representations, an example of ME-distribution which is not Phase-type, and lastly we end this section with the closure properties of the class of ME-distributions, which are the same than in the class of PH-distributions.

# 3.1 Phase-type distributions

## 3.1.1 The underlying Markov jump process

Consider a time-homogeneous Markov jump process $\{X_t\}_{t\geq 0}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a finite state space given by

$$\mathcal{E} = \{1, 2, \ldots, p, p+1\},$$

where the first $p$ states are transient and $p+1$ is absorbing.

The initial probability of $\{X_t\}_{t\geq 0}$ is denoted by the vector $(\boldsymbol{\alpha}, \alpha_{p+1})$, where

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p),$$

and

$$\alpha_i = \mathbb{P}(X_0 = i), i = 1, \ldots, p+1.$$

Let $\mathbf{Q}$ be the intensity matrix of $\{X_t\}_{t\geq 0}$. Due to the particular structure of the state space $\mathcal{E}$ and considering that every row of $\mathbf{Q}$ sums zero, $\mathbf{Q}$ admits the following block-partitioned form

$$\begin{pmatrix} \mathbf{S} & \mathbf{s} \\ \mathbf{0} & 0 \end{pmatrix},$$

where $\mathbf{S}$ is a sub-intensity $p$-square matrix formed by the $p$ transient states and $\mathbf{s} = -\mathbf{S}\mathbf{e}$, where $\mathbf{e}$ is an $p$-dimensional column vector of ones.

For every fixed transient state $i \in \mathcal{E}$, every entry of the intensity matrix $\mathbf{Q}$ satisfies

$$s_{ii} = -\sum_{j\neq i}^{p+1} s_{ij} \leq 0, \quad s_{ij} \geq 0, \forall j \neq i \quad \text{and} \quad \sum_{j=1}^{p+1} s_{ij} = 0.$$

Moreover, if we consider the transition probability matrix $\mathbf{T} = \{t_{ij}\}$ of the embedded Markov chain, then

$$s_{ij} = t_{ij}\, s_{ii} \tag{3.1}$$

for every fixed transient state $i \in \mathcal{E}$.

Based on the fact that the time spent by the Markov jump process in state $i \in \mathcal{E}$ is exponentially distributed with intensity $s_{ii}$, we have that the probability of jumping out of state $i$ in the interval $[t, t + dt)$ is $s_{ii}\, dt$. Now, if we consider that $\{X_t = i\}$, the probability of jumping to the state $j$ in the interval $[t, t+dt)$ is $t_{ij}s_{ii}\, dt$, which is equal to $s_{ij}\, dt$ by Equation (3.1).

**Lemma 3.1** *The sub-intensity matrix* $\mathbf{S}$ *is nonsingular if and only if the states* $\{1, 2, \ldots, p\}$ *are transient.*

For the proof we refer [Neu81, p. 45].

### 3.1.2 Definition of PH-distributions

Consider the stopping time

$$\tau = \inf \left\{t \geq 0 : X_t = p + 1\right\},$$

which is the time until absorption of the Markov jump process $\{X_t\}_{t\geq 0}$. Then, the random variable $\tau$ is said to be phase-type distributed (PH-distributed). The pair $(\boldsymbol{\alpha}, \mathbf{S})$ is called a representation of the distribution and we write $\tau \sim \mathrm{PH}_p\left(\boldsymbol{\alpha}, \mathbf{S}\right)$, where $p$ is called the dimension of the representation.

When $\alpha_{p+1} > 0$, it says that the distribution has an atom at zero of size $\alpha_{p+1}$ ( i.e. the distribution has point mass at zero).

Similarly than in the case of DPH-distributions, the representation of a PH-distribution is usually not unique. If the number $p$ is the smallest dimension among all the representations of the distribution of $\tau$, then $p$ is said to be the order of the representation.

**Irreducible representations of PH-distributions:**
The representation $(\boldsymbol{\alpha}, \mathbf{S})$ is called irreducible if every transient state of the underlying Markov jump process has a positive probability of being visited when the initial distribution is $\boldsymbol{\alpha}$. This is, for every transient state $i \in \mathcal{E}$

$$\left(\boldsymbol{\alpha} e^{\mathbf{S}t}\right)_i > 0,$$

for some $t > 0$.

In general, an irreducible representation may be obtained from a given reducible representation by simply removing all the superfluous states (these are states which are visited with probability zero in accordance with the initial distribution of the representation). Also, as in the case of DPH-distributions, an irreducible representation of a PH-distribution may be not unique.

Throughout this thesis, in order to simplify formulations, we assume that the PH-distributions do not have atom at zero and their representations are irreducible, unless we specify it. As well as, in many cases it will not be necessary to specified the dimension of the PH-representation.

### 3.1.3 Properties of PH-distributions

**Corollary 3.2** *Let* $X \sim \mathrm{PH}_p(\boldsymbol{\alpha}, \mathbf{S})$ *. Then, all the eigenvalues of* $\mathbf{S}$ *have strictly negative real parts.*

**Proof.** Let $\mathbf{S} = \{s_{ij}\}$ and consider $\mathcal{M} = \max_{1 \leq i \leq p} \{-s_{ii}\}$ . Then, we can write the matrix $\mathbf{S}$ as

$$\mathbf{S} = \mathcal{M}(\mathbf{T} - \mathbf{I}),$$

where $\mathbf{T}$ is a sub-transition probability matrix of an underlying Markov chain of a discrete phase-type distribution. Now, by writing

$$\mathcal{M}(\mathbf{T} - \mathbf{I}) = \mathbf{P}\mathcal{M}(\mathbf{J} - \mathbf{I})\mathbf{P}^{-1},$$

where $\mathbf{J}$ is the Jordan normal form matrix of $\mathbf{T}$, we observe that if $\lambda_i$ is an eigenvalue of $\mathbf{T}$ then the corresponding eigenvalue for $\mathbf{S}$ is

$$\mathcal{M}(\lambda_i - 1)$$

and $\mathcal{Re}(\mathcal{M}(\lambda_i - 1)) < 0$ since $\rho(\mathbf{T}) < 1$ (see Proposition 2.2). □

**Proposition 3.3** *Let* $\tau \sim \mathrm{PH}_p(\boldsymbol{\alpha}, \mathbf{S})$ *. Then,*

*i. the cumulative distribution function of* $\tau$ *is given by*

$$F_T(t) = 1 - \boldsymbol{\alpha} e^{\mathbf{S}t}\mathbf{e}, \quad t \geq 0.$$

*ii. The density function of* $\tau$ *is*

$$f_T(t) = \boldsymbol{\alpha} e^{\mathbf{S}t}\mathbf{s}, \quad t > 0 \tag{3.2}$$

*where* $\mathbf{s} = -\mathbf{S}\mathbf{e}$.

*iii. The Laplace-Stieltjes transform of* $f_T(t)$, $\mathcal{L}_\tau(s)$, *is given by*

$$\mathcal{L}_\tau(s) = \boldsymbol{\alpha}(s\mathbf{I} - \mathbf{S})^{-1}\mathbf{s}, \tag{3.3}$$

*where* $s$ *is larger than the maximum real part eigenvalue of* $\mathbf{S}$. *In particular, Equation* (3.3) *holds for all* $\mathcal{Re}(s) \geq 0$.

*iv. The expectation of $\tau$ is given by*

$$\mathbb{E}(\tau) = -\boldsymbol{\alpha}\,\mathbf{S}^{-1}\mathbf{e}.$$

*v. The $n$-th moment of $\tau$ is given by*

$$\mathbb{E}(\tau^n) = (-1)^n n!\,\boldsymbol{\alpha}\,\mathbf{S}^{-n}\mathbf{e}. \tag{3.4}$$

**Proof.** $i$. For every $i, j \in \mathcal{E}$, the transition probability of the underlying Markov jump process is given by

$$\mathbb{P}(X_t = j | X_0 = i) = \left(e^{\mathbf{Q}t}\right)_{ij},$$

where

$$e^{\mathbf{Q}t} = \begin{pmatrix} e^{\mathbf{S}t} & \mathbf{e} - e^{\mathbf{S}t}\mathbf{e} \\ \mathbf{0} & 1 \end{pmatrix},$$

see Lemma A.6. Then,

$$\begin{aligned}
\mathbb{P}(\tau > t) &= \sum_{i=1}^{p} \mathbb{P}(X_0 = i)\,\mathbb{P}(\tau > t | X_0 = i) \\
&= \sum_{i=1}^{p} \mathbb{P}(X_0 = i)\,\mathbb{P}(X_t \in \{1, 2, \ldots, p\} | X_0 = i) \\
&= \sum_{i=1}^{p} \mathbb{P}(X_0 = i) \sum_{j=1}^{p} \mathbb{P}(X_t = j | X_0 = i) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_i \left(e^{\mathbf{S}t}\right)_{ij} \\
&= \boldsymbol{\alpha}\, e^{\mathbf{S}t}\mathbf{e}.
\end{aligned}$$

$ii$. We consider to calculate $\mathbb{P}(\tau \in (t, t + dt])$.

$$\begin{aligned}
\mathbb{P}(\tau \in (t, t + dt]) &= \sum_{i=1}^{p} \sum_{j=1}^{p} \mathbb{P}(\tau \in (t, t + dt], X_t = j, X_0 = i) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{p} \mathbb{P}\left(\tau \in (t, t + dt] \middle| X_t = j, X_0 = i\right) \\
&\times\ \mathbb{P}(X_t = j | X_0 = i)\,\mathbb{P}(X_0 = i) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_i \left(e^{\mathbf{S}t}\right)_{ij} \mathbb{P}\left(\tau \in (t, t + dt] \middle| X_t = j\right) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_i \left(e^{\mathbf{S}t}\right)_{ij} s_{j,p+1}\, dt,
\end{aligned}$$

where $s_{j,p+1}dt$ is the exit rate to the absorbing state in the interval $(t, t+dt]$. Then

$$\mathbb{P}(\tau \in (t, t+dt]) = \boldsymbol{\alpha} e^{\mathbf{S}t}\mathbf{s}dt.$$

Finally, we obtain

$$f_\tau(t) = \lim_{dt\to 0} \frac{1}{dt}\mathbb{P}(\tau \in (t, t+dt]) = \boldsymbol{\alpha} e^{\mathbf{S}t}\mathbf{s}.$$

$iii.$

$$\begin{aligned}\mathcal{L}_\tau(s) &= \boldsymbol{\alpha}\int_0^\infty e^{-st}e^{\mathbf{S}t}dt\mathbf{s} \qquad \text{(recall Equation (3.2))},\\ &= \boldsymbol{\alpha}\int_0^\infty e^{-(s\mathbf{I}-\mathbf{S})t}dt\mathbf{s},\end{aligned}$$

the last integral is well-defined for all $s \in \mathbb{C}$ such that the matrix $s\mathbf{I} - \mathbf{S}$ is nonsingular (see Lemma A.3). In general, a square matrix is nonsingular if and only if it does not have a zero eigenvalue. Consider the eigenvalue of the matrix $\mathbf{S}$ with the largest real part, denoted by $\varphi_{\max}$. Then $s - \varphi_{\max}$ corresponds to the eigenvalue with the smallest real part of the matrix $s\mathbf{I} - \mathbf{S}$. Therefore, for every fixed $s \in \mathbb{C}$ such that $\mathcal{R}e(s) > \mathcal{R}e(\varphi_{max})$, all the eigenvalues of the matrix $s\mathbf{I} - \mathbf{S}$ have positive real part, which implies that $s\mathbf{I} - \mathbf{S}$ is nonsingular.

Finally, we conclude that

$$\mathcal{L}_\tau(s) = \boldsymbol{\alpha}(s\mathbf{I} - \mathbf{S})^{-1}\mathbf{s}, \quad \forall\ \mathcal{R}e(s) > \mathcal{R}e(\varphi_{max}).$$

$iv.$ This result is obtained by differentiating the Laplace-Stieltjes transform and then evaluating at $s = 0$.

$v.$ It is obtained by differentiating $n$-times the Laplace-Stieltjes transform and then evaluating at $s = 0$. □

As in the case of DPH-distributions, the entries of the matrix $-\mathbf{S}^{-1}$ have a probabilistic interpretation. In particular, $-\mathbf{S}_{ij}^{-1}$ can be interpreted as the expected time the underlying Markov jump process spent in state $j$ before it gets absorbed and starting from state $i$.

### 3.1.4 Examples

We present some examples of phase-type distributions and we start with the simplest but not less important, the Exponential distribution.

**Example 3.1 (Exponential distribution.)** *The density function of an exponential distribution with parameter* $\lambda > 0$ *is*

$$e^{-\lambda x}\lambda, \quad x \geq 0.$$

*Then, it is easy to see that a PH-representation is given by*

$$\boldsymbol{\alpha} = 1, \quad \mathbf{S} = -\lambda,$$

*where the dimension of the representation is* $1$.

The next example is also relevant ( personally I consider it important for historical reasons), since it was the first approach to work with the class of phase-type distributions.

**Example 3.2 (Erlang distribution.)** *This distribution is defined as a finite convolution of independent and identically exponential distributed random variables. Consider* $k \in \mathbb{N}$ *and* $\lambda > 0$. *Then, the density function of an Erlang distribution with parameters* $(k, \lambda)$ *is*

$$\frac{\lambda^k}{(k-1)!}x^{k-1}e^{-\lambda x}, \quad x \geq 0. \tag{3.5}$$

*A PH-representation is*

$$\boldsymbol{\alpha} = (1, 0, 0, \ldots, 0), \quad \mathbf{S} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & 0 \\ 0 & -\lambda & \lambda & \cdots & 0 \\ 0 & 0 & -\lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda \end{pmatrix},$$

*where* $\boldsymbol{\alpha}$ *is a* $k$*-dimensional row vector and* $\mathbf{S}$ *is an* $k$*-dimensional square matrix.*

**Example 3.3 (The hyper-exponential distribution.)** *This distribution is the distribution of a finite mixture of exponentially distributed random variables. Its density is given by*

$$\sum_{i=1}^{k} p_i e^{-\lambda_i x}\lambda_i, \quad x \geq 0, \tag{3.6}$$

*where* $p_i \geq 0, \lambda_i > 0$ *for all* $i = 1, \ldots, k$, *and* $\sum_{i=1}^{k} p_i = 1$. *A PH-representation for this distribution is given by*

$$\boldsymbol{\alpha} = (p_1, p_2, \ldots, p_k), \quad \mathbf{S} = \begin{pmatrix} -\lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & -\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_k \end{pmatrix}.$$

Notice that the given PH-representation for the hyper-exponential distribution is irreducible. Furthermore, if we take a different permutation of the sum in (3.6) we obtain another irreducible representation for this distribution.

**Example 3.4 (Coxian distribution.)** *A phase-type distribution with representation given by*

$$\boldsymbol{\alpha} = (p_1, p_2, p_3, \ldots, p_k), \quad \mathbf{S} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \cdots & 0 \\ 0 & 0 & -\lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda_k \end{pmatrix},$$

*is called Coxian distribution. A probabilistic interpretation of this distribution is that once the underlying Markov jump process starts in one of its transient states, let say $i \in \{1, \ldots, k\}$ then the process progresses through the subsequent states $j > i$ until it gets absorbed. This distribution was presented as an example of the class of distributions with rational Laplace-Stieltjes transform and it was an oncoming step to the class of Matrix-exponential distributions.*

### 3.1.5 Properties of the class of PH-distributions

**Theorem 3.4 (Characterization of Phase-type distributions.)** *A distribution defined on $[0, \infty)$ with rational Laplace-Stieltjes transform is of phase-type if and only if it is either the point mass at zero, or (a) it has a continuous positive density on the positive reals, and (b) its Laplace-Stieltjes transform has a unique pole of maximal real part.*

The proof is given in [O'C90b].

**Remark from Theorem 3.4.** Notice that since the poles of the Laplace-Stieltjes transform of a phase-type distribution are the eigenvalues of the sub-intensity matrix of the underlying Markov jump process. Therefore, this means that the sub-intensity matrix has a unique eigenvalue with maximal negative real part.

#### 3.1.5.1 Closure properties of the class of PH-distributions

**Theorem 3.5 (Finite convolution of phase-type distributions.)**
*Let $X \sim \mathrm{PH}_p(\boldsymbol{\alpha}, \mathbf{S})$ and $Y \sim \mathrm{PH}_q(\boldsymbol{\beta}, \mathbf{B})$ be independent random variables. Then,*

*$X+Y$ is phase-type distributed with a representation given by*

$$\left(\left(\boldsymbol{\alpha}, \mathbf{0}_{q}\right), \begin{pmatrix} \mathbf{S} & \mathbf{sb} \\ \mathbf{0}_{q\times p} & \mathbf{T} \end{pmatrix}\right),$$

*where $\mathbf{0}_{q}$ is a zero row vector of dimension $q$, $\mathbf{b} = -\mathbf{Be}$, and $\mathbf{0}_{q\times p}$ is a zero matrix dimension $q \times p$.*

The proof is given in [Neu81, p. 51]. However, we can also use the proof given in Theorem 3.13, which is merely analytical.

**Theorem 3.6 (Finite mixtures of phase-type distributions.)** *Let $\theta \in [0,1]$ and let $X$ and $Y$ be two phase-type distributed random variables with representation $(\boldsymbol{\alpha}, \mathbf{S})$ and $(\boldsymbol{\beta}, \mathbf{B})$, respectively. Then the variable defined as*

$$U = \begin{cases} X & \textit{with probability} \quad \theta, \\ Y & \textit{with probability} \quad 1-\theta, \end{cases}$$

*is phase-type distributed with a representation given by*

$$\left(\left(\theta\boldsymbol{\alpha}, (1-\theta)\boldsymbol{\beta}\right), \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}\right),$$

*where $\mathbf{0}$ is a zero matrix of appropriate dimension.*

See proof given in Theorem 3.14.

**Theorem 3.7 (Order Statistics from PH-distributions.)** *All order statistics from a finite set of independent and PH-distributed random variables are PH-distributed.*

The proof is given in Section 4.2.

**Theorem 3.8 (N-fold of PH-distributions.)**
*Consider $N \sim \mathrm{DPH}(\boldsymbol{\pi}, \mathbf{T})$ and a sequence $\{W_n\}_{n\in\mathbb{N}}$ of i.i.d. random variables with distribution $\mathrm{PH}(\boldsymbol{\beta}, \mathbf{B})$. Let $N$ be independent of the sequence $\{W_n\}_{n\in\mathbb{N}}$. Then, the random variable*

$$\sum_{n=1}^{N} W_n$$

*is phase-type distributed with a representation*

$$\left(\boldsymbol{\beta} \otimes \boldsymbol{\pi}, \mathbf{B} \otimes \mathbf{I} + \mathbf{b}\boldsymbol{\beta} \otimes \mathbf{T}\right),$$

*where $\mathbf{b} = -\mathbf{Be}$.*

The proof is found in [Neu81, p. 53].

## 3.2 Matrix-Exponential distributions

**Definition 3.9** A non-negative random variable $X$ is said to be Matrix-Exponential distributed (ME-distributed) if its density function, $f_X(x)$, can be written in the form

$$f_X(x) = \boldsymbol{\alpha} e^{\mathbf{S}x}\mathbf{s}, \quad x > 0.$$

where $\boldsymbol{\alpha}$ is an $p$-dimensional row vector, $\mathbf{S}$ is a $p$-square matrix, $\mathbf{s}$ is a $p$-dimensional column vector, and $\boldsymbol{\alpha}$, $\mathbf{S}$ and $\mathbf{s}$ can have complex entries. We write $X \sim \mathrm{ME}_p(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$, where $(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$ is called the ME-representation of the distribution of $X$ and the value $p$ is the dimension of the representation (also called the order of the representation).

If $1 - \int_0^\infty \boldsymbol{\alpha} e^{\mathbf{S}x}\mathbf{s}dx = \alpha_{p+1} > 0$, then the distribution of $X$ is said to have an atom at zero of size $\alpha_{p+1}$.

In order to simplify formulations, throughout this thesis we assume that all Matrix-Exponential distributed random variables do not have an atom at zero.

### 3.2.1 Properties of ME-distributions

**Theorem 3.10** *Let $X \sim \mathrm{ME}_p(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$ such that $\mathbf{S}$ is nonsingular. Then*

*i. The cumulative distribution function of $X$ is given by*

$$F_X(x) = 1 + \boldsymbol{\alpha} e^{\mathbf{S}x}\mathbf{S}^{-1}\mathbf{s}.$$

*ii. The Laplace-Stieltjes transform is given by*

$$\mathcal{L}_X(s) = \boldsymbol{\alpha}(s\mathbf{I} - \mathbf{S})^{-1}\mathbf{s},$$

*where $s$ is larger than the maximum eigenvalue of $\mathbf{S}$.*

*iii. The $n$-th moment of $X$ by*

$$\mathbb{E}(X^n) = (-1)^{n+1} n! \boldsymbol{\alpha}(\mathbf{S})^{-(n+1)}\mathbf{s}, \quad n \in \mathbb{N}.$$

**Proof.** $i.$

$$F_X(x) = \boldsymbol{\alpha}\int_0^x e^{\mathbf{S}t}dt\mathbf{s} \quad = \quad \boldsymbol{\alpha} e^{\mathbf{S}x}\mathbf{S}^{-1}\mathbf{s} - \boldsymbol{\alpha}\mathbf{S}^{-1}\mathbf{s}, \quad \text{(see Lemma A.3)}$$

where $-\boldsymbol{\alpha}\mathbf{S}^{-1}\mathbf{s} = \int_0^\infty f_X(t)dt = 1$.

The proof for $ii.$ and $iii.$ is exactly the same than the proof in statements $iii.$ and $v.$ of Proposition 3.3. □

### 3.2.2 Properties of ME-representations

**Theorem 3.11** *Let* $X \sim \mathrm{ME}(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$. *Then, there exists another ME-representation* $(\boldsymbol{\varsigma}, \mathbf{Z}, \mathbf{z})$ *for the distribution of* $X$ *such that all the eigenvalues of* $\mathbf{Z}$ *have negative real part.*

See [BN17, p. 203] for the proof.

Theorem 3.11 implies that for every ME-distributed random variable we can find a ME-representation $(\boldsymbol{\varsigma}, \mathbf{Z}, \mathbf{z})$ such that the matrix $\mathbf{Z}$ is nonsingular.

In [AB95] is shown that the class of distributions with rational Laplace-Stieltjes transform is the same than the class of Matrix-Exponential distributions. Furthermore, it is proved that given a rational Laplace-Stieltjes transform of a ME-distributed random variable we obtain a ME-representation. The criterion for finding the ME-representation is explained in the following theorem.

**Theorem 3.12 (The companion form representation of ME-distributions.)** *Let* $X$ *be ME-distributed with Laplace-Stieltjes transform given by*

$$\mathcal{L}_X(s) = \frac{b_p s^{p-1} + \cdots + b_{p-1}s + b_1}{s^p + a_p s^{p-1} + \cdots + a_2 s + a_1}$$

*where* $a_1, \ldots, a_p, b_1, \ldots, b_p$ *are real values. Then* $X$ *has a representation* $(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$, *where*

$$\begin{aligned}
\boldsymbol{\alpha} &= (b_1, b_2, \ldots, \ldots, b_p), \\
\mathbf{S} &= \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ -a_1 & -a_2 & -a_3 & -a_4 & \cdots & -a_p \end{pmatrix} \quad and \quad \mathbf{s} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.
\end{aligned}$$

*The matrix* $\mathbf{S}$ *is called the companion matrix.*

### 3.2.3 Examples

**Example 3.5** *Consider the density function given by*

$$f(t) = 2e^{-t}\left(1 - cos(t)\right), \quad t > 0,$$

*where its the Laplace transform is given by*

$$\mathcal{L}(s) = \frac{2}{s^3 + 3s^2 + 4s + 2}\,.$$

*Then, $f(t)$ is a density of a ME-distributed random variable. By using the companion form representation (see Theorem 3.12), we obtain a ME-representation $(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$, where*

$$\begin{aligned} \boldsymbol{\alpha} &= (2,0,0) \\ \mathbf{S} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -4 & -3 \end{pmatrix} \quad \textit{and} \quad \mathbf{s} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \end{aligned}$$

*Also, this density function can not be of phase-type since in the set $\{2n\pi : n \in \mathbb{N}\}$ the function $f(t)$ is equal to $0$ (see Theorem 3.4).*

### 3.2.4 Closure properties of the class of ME-distributions

Matrix-exponential distribution are closed under finite convolutions, finite mixtures and order statistics.

**Theorem 3.13 (Finite convolution of ME-distributions.)** *Let $X \sim \mathrm{ME}_p(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$ and $Y \sim \mathrm{ME}_q(\boldsymbol{\beta}, \mathbf{B}, \mathbf{b})$ be two independent random variables. Then, $X+Y$ is Matrix-exponentially distributed with a representation given by $(\boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta})$, where*

$$\boldsymbol{\sigma} = \left(\boldsymbol{\alpha}, \mathbf{0}_{q}\right), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{S} & \mathbf{s}\boldsymbol{\beta} \\ \mathbf{0}_{q\times p} & \mathbf{B} \end{pmatrix}, \quad \boldsymbol{\eta} = \left(\mathbf{0}_{p}, \mathbf{b}\right)^{\top},$$

*and $\mathbf{0}_q, \mathbf{0}_p$ are zero row vectors of dimension $q$ and $p$, respectively, while $\mathbf{0}_{q\times p}$ is an $q \times p$-dimensional zero matrix.*

**Proof.** Notice that since $X$ and $Y$ are independent, then the Laplace transform of the distribution of the sum $X+Y$ is given by the product of the Laplace transfor of the distribution of $X$ and $Y$, this is

$$\mathcal{L}_{X+Y}(s) = \mathcal{L}_X(s)\mathcal{L}_Y(s).$$

That shows that the Laplace transform of $X+Y$ is a rational function and consequently $X+Y$ is a Matrix-exponential distributed random variable.

In the next, we are going to prove that $(\boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta})$ is in fact a representation for the distribution of $X+Y$. By calculating the inverse of the matrix $s\mathbf{I}-\boldsymbol{\Sigma}$ (see Lemma A.4), we obtain

$$\begin{aligned}(s\mathbf{I}-\boldsymbol{\Sigma})^{-1} &= \begin{pmatrix} s\mathbf{I}-\mathbf{S} & -\mathbf{s}\boldsymbol{\beta} \\ \mathbf{0}_{q\times p} & s\mathbf{I}-\mathbf{B} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (s\mathbf{I}-\mathbf{S})^{-1} & (s\mathbf{I}-\mathbf{S})^{-1}\mathbf{s}\boldsymbol{\beta}(s\mathbf{I}-\mathbf{B})^{-1} \\ \mathbf{0}_{q\times p} & (s\mathbf{I}-\mathbf{B})^{-1} \end{pmatrix}.\end{aligned}$$

Then,

$$\begin{aligned}\boldsymbol{\sigma}(s\mathbf{I}-\boldsymbol{\Sigma})^{-1}\boldsymbol{\eta} &= \boldsymbol{\alpha}(s\mathbf{I}-\mathbf{S})^{-1}\mathbf{s}\boldsymbol{\beta}(s\mathbf{I}-\mathbf{B})^{-1}\mathbf{b} \\ &= \mathcal{L}_X(s)\mathcal{L}_Y(s).\end{aligned}$$

Therefore, $(\boldsymbol{\sigma}, \boldsymbol{\Sigma}, \boldsymbol{\eta})$ is a representation for the distribution of $X+Y$. □

**Theorem 3.14 (Finite mixtures of Matrix-exponential distributions.)** *Let $\theta \in [0,1]$ and let $X$ and $Y$ be two independent Matrix-exponential distributed random variables with representation $(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$ and $(\boldsymbol{\beta}, \mathbf{B}, \mathbf{b})$, respectively. Then, the random variable defined as*

$$U = \begin{cases} X & \text{with probability } \theta, \\ Y & \text{with probability } 1-\theta, \end{cases}$$

*is Matrix-exponential distributed with a representation given by*

$$\left((\theta\boldsymbol{\alpha}, (1-\theta)\boldsymbol{\beta}), \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}, \begin{pmatrix} \mathbf{s} \\ \mathbf{b} \end{pmatrix}\right),$$

*where $\mathbf{0}$ represents a zero matrix of appropriate dimension.*

**Proof.** The random variable $U$ can be written as $\mathbf{1}_{\{\mathcal{A}\}}X + \mathbf{1}_{\{\mathcal{A}^c\}}Y$, where $\mathbb{P}(\mathcal{A}) = \theta$ and $\mathbb{P}(\mathcal{A}^c) = 1-\theta$. Then, the Laplace transform of the distribution of $U$ is calculated as follows

$$\begin{aligned}\mathcal{L}_U(s) = \mathbb{E}\left(e^{-sU}\right) &= \mathbb{P}(\mathcal{A})\,\mathbb{E}\left(e^{-sU}|\mathcal{A}\right) + \mathbb{P}(\mathcal{A}^c)\,\mathbb{E}\left(e^{-sU}|\mathcal{A}^c\right) \\ &= \theta\mathbb{E}\left(e^{-sX}\right) + (1-\theta)\mathbb{E}\left(e^{-sY}\right) \\ &= \theta\boldsymbol{\alpha}(s\mathbf{I}-\mathbf{S})^{-1}\mathbf{s} + (1-\theta)\boldsymbol{\beta}(s\mathbf{I}-\mathbf{B})^{-1}\mathbf{b} \\ &= (\theta\boldsymbol{\alpha}, (1-\theta)\boldsymbol{\beta})\begin{pmatrix} s\mathbf{I}-\mathbf{S} & \mathbf{0} \\ \mathbf{0} & s\mathbf{I}-\mathbf{B} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{s} \\ \mathbf{b} \end{pmatrix} \\ &= (\theta\boldsymbol{\alpha}, (1-\theta)\boldsymbol{\beta})\left(s\mathbf{I} - \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}\right)^{-1}\begin{pmatrix} \mathbf{s} \\ \mathbf{b} \end{pmatrix}.\end{aligned}$$

□

**Theorem 3.15 (Order Statistics from ME-distributions.)** *All order statistics from a set of independent and ME-distributed random variables are ME-distributed.*

We present the proof in Section 4.3.

**Theorem 3.16 (N-fold of ME-distributions.)** *Consider* $N \sim \mathrm{MG}(\boldsymbol{\beta}, \mathbf{B}, \mathbf{b})$ *and a sequence* $\{W_n\}_{n\in\mathbb{N}}$ *of i.i.d. random variables with distribution* $\mathrm{ME}(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$, $n \in \mathbb{N}$. *Let* $N$ *be independent of the sequence* $\{W_n\}_{n\in\mathbb{N}}$. *Then, the random variable*

$$W = \sum_{n=1}^{N} W_n$$

*is Matrix-exponential distributed with a representation given by*

$$(\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \mathbf{I} \otimes \mathbf{S} + \mathbf{s}\boldsymbol{\alpha} \otimes \mathbf{B}, \mathbf{s} \otimes \mathbf{b}).$$

The exact proof given in [Neu81, p. 53] is applied for the proof of this theorem.

CHAPTER 4

# Order Statistics from Matrix-Exponential distributions

In this chapter we consider order statistics from independent and non-identically distributed random variables and our aim is to obtain representations for the corresponding case according to the distribution, this is when the distribution is discrete Phase-type distribution, Matrix-geometric distribution, Phase-type distribution and Matrix-exponential distribution. The chapter contains five main sections which are explained next.

We start with a introductory section about order statistics and their formula of the distribution function in the general case.

The second and the third section recalls PH-representations and ME-representations, respectively, which are already known for the corresponding case of order statistics from PH-distributions and from ME-distributions.

In the fourth section we present DPH-representations that we have proposed for order statistics from discrete phase-type distributions. Finally, we finish this chapter with a section (the fifth section) where we provide two types of representations for the distribution of order statistics from matrix-geometric distributions.

To simplify calculations we assume that none of the random variables have an atom at zero ( or a point at zero in the case of discrete time).

## 4.1 Introduction to Order Statistics

In the literature we can find a formula for the distribution of order statistics from independent and non-identically distributed random variables in terms of the so-called permanents. Here we recall the formula for the survival function of order statistics, which is derived from the formula based on permanents.

Let $\mathbf{A} = \{a_{ij}\}$ be an $n$-square matrix and let $\mathcal{P}_n$ denote the set of all the permutations of the indices $\{1, 2, \dots, n\}$. Then the permanent of $\mathbf{A}$ is defined by

$$\text{per}\big[\mathbf{A}\big] = \sum_{\sigma \in \mathcal{P}_n} \prod_{i=1}^{n} a_{i,\sigma(i)}. \tag{4.1}$$

Over the sum of the permanent formula is considered the exact same permutations as the determinant formula, however the difference is that the permanent formula does not consider negative signs over the sum.

Let $Y_1, Y_2, \dots, Y_n$ be $n$ independent random variables with distribution function $F_j(y)$, $j = 1, \dots, n$, respectively. Let $Y_{(1:n)} \leq Y_{(2:n)} \leq \cdots \leq Y_{(n:n)}$ denote the corresponding order statistics. Then, the survival function of $Y_{(r:n)}$ can be expressed as

$$\mathbb{P}\left(Y_{(r:n)} > y\right) = \sum_{i=0}^{r-1} \frac{1}{i!(n-i)!} \text{per}\Big[\underbrace{\mathbf{F}(y)}_{\substack{i \\ \text{columns}}} \quad \underbrace{\mathbf{e} - \mathbf{F}(y)}_{\substack{n-i \\ \text{columns}}}\Big], \quad y \in \mathbb{R}, \tag{4.2}$$

where $r = 1, \dots, n$, $\mathbf{F}(y) = \big(F_1(y), F_2(y), \dots, F_n(y)\big)^{\top}$ and $\underbrace{\mathbf{F}(y)}_{\substack{i \\ \text{columns}}}$ represents $i$ repetitions of the column vector $\mathbf{F}(y)$ (see [BB89]).

## 4.2 Order Statistics from Phase-type distributions

Consider $n$ independent random variables $Y_1, \dots, Y_n$ which are phase-type distributed $Y_i \sim \text{PH}\left(\boldsymbol{\alpha}_i, \mathbf{S}_i\right), i = 1, \dots, n$.

In this section we are going to find PH-representations for the order statistics $Y_{(k:n)}$, $k = 1, \ldots, n$, based on a probabilistic interpretation. For our aim we are going to introduce some notation that allow us to express the representations of order statistics in a general way and we give the corresponding probabilistic interpretation of them. The idea of these representations was originally given by M. F. Neuts in [Neu81, p. 60].

Let $\left\{X_t^i\right\}_{t\geq 0}$ denote the underlying Markov jump process of the distribution of the random variable $Y_i$ and let $\mathcal{E}_i = \{1, 2, \ldots, p_i, p_i + 1\}$ be its state space.

Consider the multivariate Markov jump process formed by

$$\mathbf{X}_t = \left(X_t^1, \ldots, X_t^n\right),$$

whose state space is given by $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2 \times \cdots \times \mathcal{E}_n$. We consider a lexicographical ordering for the state space $\mathcal{E}$. Then, the initial distribution of $\mathbf{X}_t$ can be expressed by

$$\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n$$

and its sub-intensity matrix by

$$\mathbf{S}_1 \oplus \cdots \oplus \mathbf{S}_n.$$

Observe that matrix $\mathbf{S}_1 \oplus \cdots \oplus \mathbf{S}_n$ covers only the rates for transitions between transient states, this represents the case when every Markov jump process is in a transient state (or none of the Markov jump process are not in the absorbing state).

The $k$-th order statistic $Y_{(k:n)}$ is interpreted as the time when the $k$-th Markov jump process gets absorbed in the multivariable Markov jump process. By following this idea, we can easily get a PH-representation for the distribution of the first order statistic $Y_{(1:n)}$, which is given by

$$\left(\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n, \mathbf{S}_1 \oplus \cdots \oplus \mathbf{S}_n\right).$$

Now, in order to obtain a representation for the distribution of the second order statistic $Y_{(2:n)}$, we need to take into account that one Markov jump process has been absorbed already and $n-1$ Markov jump processes keep being in their corresponding set of transient states. All the intensity rates corresponding to those possibilities are contained in the following matrix

$$\mathbf{G}_{2:n} = \begin{pmatrix} \mathbf{S}_1 \oplus \cdots \oplus \mathbf{S}_n & \mathbf{I} \otimes \cdots \otimes \mathbf{I} \otimes \mathbf{s}_n & \mathbf{I} \otimes \cdots \otimes \mathbf{s}_{n-1} \otimes \mathbf{I} & \cdots & \mathbf{s}_1 \otimes \mathbf{I} \cdots \otimes \mathbf{I} \\ \mathbf{0} & \mathbf{S}_1 \oplus \cdots \oplus \mathbf{S}_{n-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_1 \oplus \cdots \oplus \mathbf{S}_{n-2} \oplus \mathbf{S}_n & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_2 \oplus \cdots \oplus \mathbf{S}_n \end{pmatrix}$$

where $\mathbf{s}_i = -\mathbf{S}_i\,\mathbf{e}, i = 1, \ldots, n$. Thus, a PH-representation for the second order statistic $Y_{(2:n)}$ is

$$((\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n, \mathbf{0}), \mathbf{G}_{2:n}).$$

In order to generalise the last idea on how to obtain PH-representations for the rest of the order statistics, we are going to consider the case $n = 3$ and calculate the representation for the maximum $Y_{(3:3)}$. Thus, we consider the case when a second Markov jump process gets absorbed. The corresponding transition rates are enclosed in the following matrix

$$\mathbf{G}_{3:3} = \begin{pmatrix} \mathbf{S}_1 \oplus \mathbf{S}_2 \oplus \mathbf{S}_3 & \mathbf{I} \otimes \mathbf{I} \otimes \mathbf{s}_3 & \mathbf{I} \otimes \mathbf{s}_2 \otimes \mathbf{I} & \mathbf{s}_1 \otimes \mathbf{I} \otimes \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_1 \oplus \mathbf{S}_2 & \mathbf{0} & \mathbf{0} & \mathbf{I} \otimes \mathbf{s}_2 & \mathbf{s}_1 \otimes \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_1 \oplus \mathbf{S}_3 & \mathbf{0} & \mathbf{I} \otimes \mathbf{s}_3 & \mathbf{0} & \mathbf{s}_1 \otimes \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{S}_2 \oplus \mathbf{S}_3 & \mathbf{0} & \mathbf{I} \otimes \mathbf{s}_3 & \mathbf{s}_2 \otimes \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{S}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{S}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{S}_3 \end{pmatrix}.$$

Then, a PH-representation for the maximum $Y_{(3:3)}$ is

$$((\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n, \mathbf{0}, \mathbf{0}), \mathbf{G}_{3:3}).$$

The matrix $\mathbf{G}_{3:3}$ can be expressed as

$$\mathbf{G}_{3:3} = \begin{pmatrix} \mathbf{S}_{(0,0)} & \mathbf{S}_{(0,1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{(1,1)} & \mathbf{S}_{(1,2)} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{(2,2)} \end{pmatrix}, \tag{4.3}$$

where

$$\begin{aligned} \mathbf{S}_{(0,0)} &= \mathbf{S}_1 \oplus \mathbf{S}_2 \oplus \mathbf{S}_3, \\ \mathbf{S}_{(0,1)} &= \begin{pmatrix} \mathbf{I} \otimes \mathbf{I} \otimes \mathbf{s}_3 & \mathbf{I} \otimes \mathbf{s}_2 \otimes \mathbf{I} & \mathbf{s}_1 \otimes \mathbf{I} \otimes \mathbf{I} \end{pmatrix}, \\ \mathbf{S}_{(1,1)} &= \begin{pmatrix} \mathbf{S}_1 \oplus \mathbf{S}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_1 \oplus \mathbf{S}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_2 \oplus \mathbf{S}_3 \end{pmatrix}, \end{aligned}$$

$$\mathbf{S}_{(1,2)} = \begin{pmatrix} \mathbf{I}\otimes\mathbf{s}_2 & \mathbf{s}_1\otimes\mathbf{I} & \mathbf{0} \\ \mathbf{I}\otimes\mathbf{s}_3 & \mathbf{0} & \mathbf{s}_1\otimes\mathbf{I} \\ \mathbf{0} & \mathbf{I}\otimes\mathbf{s}_3 & \mathbf{s}_2\otimes\mathbf{I} \end{pmatrix},$$

$$\mathbf{S}_{(2,2)} = \begin{pmatrix} \mathbf{S}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_3 \end{pmatrix}.$$

Our purpose is to show a compact form for a representation of the $k$-th order statistic, similar to the matrix $\mathbf{G}_{3:3}$ in Equation (4.3).

Since the $k$-th order statistic is obtained as the first time when $k$ process have became absorbed, the sub-intensity matrix of the $k$-th order statistic encloses the rates of the transitions between states of the multivariate Markov jump process which have less than $k$ absorbing states in their entries (which means that there are less than $k$ Markov jump processes that have not became absorbed).

For a fixed $j = 1, 2, \ldots, n$, consider the set of all combinations given by $j$ elements of the set $\{1, 2, \ldots, n\}$ and we write every combination in a vector form. We place all the vectors of combinations in the following set

$$\mathcal{C}_{j,n} = \left\{ \left(c^\theta_1, c^\theta_2, \ldots, c^\theta_j\right) : c^\theta_i \in \{1, 2, \ldots, n\},\ \ \theta \in \{1, \ldots, \tfrac{n!}{j!(n-j)!}\} \right\} \tag{4.4}$$

and we set a lexicographical ordering in $\mathcal{C}_{j,n}$. We call $\left(c^\theta_1, c^\theta_2, \ldots, c^\theta_j\right)$ as an $j$-tuple and $\boldsymbol{\theta}$ denotes the $j$-tuple given by $\left(c^\theta_1, c^\theta_2, \ldots, c^\theta_j\right)$. We refer to $\boldsymbol{\theta}$ as the $\boldsymbol{\theta}$-tuple.

For $k = 0, \ldots, n-1$, define the block diagonal matrix

$$\mathbf{S}_{(k,k)} = \text{diag}\left\{ \left(\mathbf{S}_{c^l_1} \oplus \cdots \oplus \mathbf{S}_{c^l_{n-k}}\right) : \quad \left(c^l_1, c^l_2, \ldots, c^l_{n-k}\right) \in \mathcal{C}_{n-k,n} \right\}. \tag{4.5}$$

Now, for $k = 1, \ldots, n-1$, define the matrix

$$\mathbf{S}_{(k-1,k)} = \begin{pmatrix} \mathbf{g}(\boldsymbol{\ell}_1, \boldsymbol{\mu}_1) & \mathbf{g}(\boldsymbol{\ell}_1, \boldsymbol{\mu}_2) & \cdots & \mathbf{g}(\boldsymbol{\ell}_1, \boldsymbol{\mu}_{n-k}) \\ \mathbf{g}(\boldsymbol{\ell}_2, \boldsymbol{\mu}_1) & \mathbf{g}(\boldsymbol{\ell}_2, \boldsymbol{\mu}_2) & \cdots & \mathbf{g}(\boldsymbol{\ell}_2, \boldsymbol{\mu}_{n-k}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{g}(\boldsymbol{\ell}_{n-k+1}, \boldsymbol{\mu}_1) & \mathbf{g}(\boldsymbol{\ell}_{n-k+1}, \boldsymbol{\mu}_2) & \cdots & \mathbf{g}(\boldsymbol{\ell}_{n-k+1}, \boldsymbol{\mu}_{n-k}) \end{pmatrix}, \tag{4.6}$$

where

$$\mathbf{g}(\boldsymbol{\ell}, \boldsymbol{\mu}) = \begin{cases} \mathbf{0}_{|\boldsymbol{\ell}|\times|\boldsymbol{\mu}|} & \text{if} \quad \boldsymbol{\ell}\cap\boldsymbol{\mu} = \emptyset \text{ or } \boldsymbol{\ell}\cap\boldsymbol{\mu} \neq \boldsymbol{\mu}, \\ \mathbf{I}\otimes\cdots\otimes\mathbf{I}\otimes\mathbf{s}_{c*}\otimes\mathbf{I}\otimes\cdots\otimes\mathbf{I} & \text{if} \quad \boldsymbol{\ell}\cap\boldsymbol{\mu} = \boldsymbol{\mu} \text{ and } \boldsymbol{\ell}-\boldsymbol{\mu} = c*, \end{cases}$$

where each $\boldsymbol{\ell}$ and $\boldsymbol{\mu}$ represents a combination in a vector form and $\mathbf{s}_{c*}$ is located according to the position of the number $c*$ in the $\boldsymbol{\ell}$-tuple.

Finally we present the corresponding representations for the order statistics of phase-type distributed random variables.

**Theorem 4.1** *Let $Y_j \sim \mathrm{PH}(\boldsymbol{\alpha}_j, \mathbf{S}_j)$, $j = 1, \ldots, n$, be $n$ independent random variables. Then, the $k$-th order statistic, $Y_{(k:n)}$, has a PH-representation $\left(\boldsymbol{\alpha}_{(k:n)}, \mathbf{G}_{(k:n)}\right)$, where*

$$\boldsymbol{\alpha}_{k:n} = \left(\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n, \mathbf{0}, \ldots, \mathbf{0}, \mathbf{0}\right), \tag{4.7}$$

$$\mathbf{G}_{k:n} = \begin{pmatrix} \mathbf{S}_{(0,0)} & \mathbf{S}_{(0,1)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{(1,1)} & \mathbf{S}_{(1,2)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots\vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_{(k-2,k-2)} & \mathbf{S}_{(k-2,k-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{S}_{(k-1,k-1)} \end{pmatrix}, \tag{4.8}$$

*for every $k = 1, 2, \ldots, n$.*

**Proof.** The proof is based on a probabilistic interpretation. Consider the multivariate Markov jump process formed with the underlying Markov jump processes of the phase-type distributed random variables $Y_j, j = 1, \ldots, n$. That is denoted by $\left(X_t^1, \ldots, X_t^n\right)$. Since the $k$-th order statistic $Y_{(k:n)}$ is interpreted as the first time when $k$ processes have became absorbed in the multivariate Markov jump process $\left(X_t^1, \ldots, X_t^n\right)$, then the sub-intensity matrix of the representation for $Y_{(k:n)}$ only needs to take into account the intensities of the transitions up to $n - k + 1$ Markov jump processes keep being in a transient state. $\square$

**Corollary 4.2 (Minimum and Maximum)** *Let $Y_j \sim \mathrm{PH}(\boldsymbol{\alpha}_j, \mathbf{S}_j)$, $j = 1, \ldots, n$, be $n$ independent random variables. Then, a PH-representation for the minimum is*

$$Y_{(1:n)} \sim \mathrm{PH}\left(\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n, \mathbf{S}_1 \oplus \cdots \oplus \mathbf{S}_n\right), \tag{4.9}$$

*and a PH-representation for the maximum is*

$$Y_{(n:n)} \sim \mathrm{PH}\left(\boldsymbol{\alpha}_{n:n}, \mathbf{G}_{n:n}\right). \tag{4.10}$$

*Recall Theorem 4.1.*

## 4.3 Order Statistics from Matrix-exponential distributions

In this section we prove that we can use the same form of the representation of the $k$-th order statistic given in Theorem 4.1 for the representation of the $k$-th order statistic in the case of matrix-exponential distributions. The proof is done analytically and to help to understand the calculations we have broken the proof into two parts. We present the first part in the following Lemma.

**Lemma 4.3** *Let $Y_1, \ldots, Y_n$ be $n$ independent and ME-distributed random variables with $Y_j \sim \mathrm{ME}(\boldsymbol{\alpha}_j, \mathbf{S}_j, \mathbf{s}_j)$ and $\mathbf{s}_j = -\mathbf{S}_j \mathbf{e}$ for all $j = 1, \ldots, n$. Consider the row vector $\boldsymbol{\alpha}_{k:n}$ given in Equation (4.7) and the block matrix $\mathbf{G}_{k:n}$ given in Equation (4.8) which are now constructed with the matrix-exponential representations $\mathrm{ME}(\boldsymbol{\alpha}_j, \mathbf{S}_j, \mathbf{s}_j)$, $j = 1, \ldots, n$. Then, for every fixed $k = 1, \ldots, n$, we have*

$$\boldsymbol{\alpha}_{k:n}\, e^{\mathbf{G}_{k:n} y} = \begin{pmatrix} \mathbf{B}_{(0,0)} & \mathbf{B}_{(0,1)} & \cdots & \mathbf{B}_{(0,k-1)} \end{pmatrix}, \tag{4.11}$$

*where*

$$\mathbf{B}_{(0,k)} = \left( \left( \underset{u \in \boldsymbol{\ell}}{\otimes} \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \right) \left( \prod_{v \in \boldsymbol{\ell}^c} F_v(y) \right) \right)_{\boldsymbol{\ell} \in \mathcal{C}_{n-k,n}}$$

*and where $\boldsymbol{\ell}^c$ is the complement set of $\boldsymbol{\ell}$.*

**Proof.** Let us abbreviate $\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n$ by $\overline{\boldsymbol{\alpha}}_n$. Let $y > 0$ and take $k = 2$. Then, we have

$$\begin{aligned} \boldsymbol{\alpha}_{2:n} e^{\mathbf{G}_{2:n} y} &= (\overline{\boldsymbol{\alpha}}_n, \mathbf{0}) \begin{pmatrix} e^{\mathbf{S}_{(0,0)} y} & \int_0^y e^{\mathbf{S}_{(0,0)}(y-t)} \mathbf{S}_{(0,1)} e^{\mathbf{S}_{(1,1)} t} dt \\ \mathbf{0} & e^{\mathbf{S}_{(1,1)} y} \end{pmatrix} \\ &= \left( \overline{\boldsymbol{\alpha}}_n e^{\mathbf{S}_{(0,0)} y}, \overline{\boldsymbol{\alpha}}_n \int_0^y e^{\mathbf{S}_{(0,0)}(y-t)} \mathbf{S}_{(0,1)} e^{\mathbf{S}_{(1,1)} t} dt \right). \end{aligned}$$

Now we focus on calculating the integral

$$\int_0^y \overline{\boldsymbol{\alpha}}_n e^{\mathbf{S}_{(0,0)}(y-t)} \mathbf{S}_{(0,1)} e^{\mathbf{S}_{(1,1)} t} dt. \tag{4.12}$$

First observe that

$$\overline{\boldsymbol{\alpha}}_n e^{\mathbf{S}_{(0,0)}(y-t)} = \left( \boldsymbol{\alpha}_1 e^{\mathbf{S}_1 (y-t)} \right) \otimes \cdots \otimes \left( \boldsymbol{\alpha}_n e^{\mathbf{S}_n (y-t)} \right), \tag{4.13}$$

then we multiply Equation (4.13) by $\mathbf{S}_{(0,1)}$, which results to

$$\overline{\boldsymbol{\alpha}}_n e^{\mathbf{S}_{(0,0)}(y-t)} \mathbf{S}_{(0,1)} =$$

$$\begin{aligned}\Big(&\boldsymbol{\alpha}_1 e^{\mathbf{S}_1(y-t)} \otimes \cdots \otimes \boldsymbol{\alpha}_{n-1} e^{\mathbf{S}_{n-1}(y-t)} \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_n(y-t)} \mathbf{s}_n, \\ &\boldsymbol{\alpha}_1 e^{\mathbf{S}_1(y-t)} \otimes \cdots \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_{n-1}(y-t)} \mathbf{s}_{n-1} \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_n(y-t)}, \\ &\quad\vdots \\ &\boldsymbol{\alpha}_1 e^{\mathbf{S}_1(y-t)} \mathbf{s}_1 \otimes \boldsymbol{\alpha}_2 e^{\mathbf{S}_2(y-t)} \otimes \cdots \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_n(y-t)}\Big).\end{aligned} \tag{4.14}$$

Next, we multiply the last product by $e^{\mathbf{S}_{(1,1)}t}$,

$$\overline{\boldsymbol{\alpha}}_n\, e^{\mathbf{S}_{(0,0)}(y-t)} \mathbf{S}_{(0,1)} e^{\mathbf{S}_{(1,1)}t} =$$

$$\begin{aligned}\Big(&\boldsymbol{\alpha}_1 e^{\mathbf{S}_1 y} \otimes \cdots \otimes \boldsymbol{\alpha}_{n-1} e^{\mathbf{S}_{n-1} y} \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_n(y-t)} \mathbf{s}_n, \\ &\boldsymbol{\alpha}_1 e^{\mathbf{S}_1 y} \otimes \cdots \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_{n-1}(y-t)} \mathbf{s}_{n-1} \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_n y}, \\ &\quad\vdots \\ &\boldsymbol{\alpha}_1 e^{\mathbf{S}_1(y-t)} \mathbf{s}_1 \otimes \boldsymbol{\alpha}_2 e^{\mathbf{S}_2 y} \otimes \ldots \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_n y}\Big).\end{aligned} \tag{4.15}$$

Lastly, by integrating, we obtain the next row vector

$$\int_0^y \overline{\boldsymbol{\alpha}}_n\, e^{\mathbf{S}_{(0,0)}(y-t)} \mathbf{S}_{(0,1)} e^{\mathbf{S}_{(1,1)}t} dt =$$

$$\begin{aligned}\Big(&\boldsymbol{\alpha}_1 e^{\mathbf{S}_1 y} \otimes \cdots \otimes \boldsymbol{\alpha}_{n-1} e^{\mathbf{S}_{n-1} y} \otimes F_n(y), \\ &\boldsymbol{\alpha}_1 e^{\mathbf{S}_1 y} \otimes \cdots \otimes F_{n-1}(y) \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_n y}, \\ &\quad\vdots \\ &F_1(y) \otimes \boldsymbol{\alpha}_2 e^{\mathbf{S}_2 y} \otimes \ldots \otimes \boldsymbol{\alpha}_n e^{\mathbf{S}_n y}\Big).\end{aligned}$$

Therefore, we can write

$$\int_0^y \overline{\boldsymbol{\alpha}}_n\, e^{\mathbf{S}_{(0,0)}(y-t)} \mathbf{S}_{(0,1)} e^{\mathbf{S}_{(1,1)}t} dt = \left( \left( \bigotimes_{u \in \boldsymbol{\ell}} \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \right) \left( \prod_{v \in \boldsymbol{\ell}^c} F_v(y) \right) \right)_{\boldsymbol{\ell} \in \mathcal{C}_{n-1,n}} . \tag{4.16}$$

Now, if we denote

$$\mathbf{B}_{(0,0)} = \overline{\boldsymbol{\alpha}}_n\, e^{\mathbf{S}_{(0,0)} y} \quad \text{and} \quad \mathbf{B}_{(0,1)} = \left( \left( \bigotimes_{u \in \boldsymbol{\ell}} \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \right) \left( \prod_{v \in \boldsymbol{\ell}^c} F_v(y) \right) \right)_{\boldsymbol{\ell} \in \mathcal{C}_{n-1,n}} .$$

Then,

$$\boldsymbol{\alpha}_{2:n} e^{\mathbf{G}_{2:n} y} = \begin{pmatrix} \mathbf{B}_{(0,0)} & \mathbf{B}_{(0,1)} \end{pmatrix}.$$

Our objective is to generalize the last idea. In fact, for every $k = 1, \ldots, n$, we can write

$$\boldsymbol{\alpha}_{k:n} e^{\mathbf{G}_{k:n} y} = \begin{pmatrix} \mathbf{B}_{(0,0)} & \mathbf{B}_{(0,1)} & \cdots & \mathbf{B}_{(0,k-1)} \end{pmatrix}. \tag{4.17}$$

Thus, we only need to show that each block entry is given by

$$\mathbf{B}_{(0,k)} = \left( \left( \bigotimes_{u \in \boldsymbol{\ell}} \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \right) \left( \prod_{v \in \boldsymbol{\ell}^c} F_v(y) \right) \right)_{\boldsymbol{\ell} \in \mathcal{C}_{n-k,n}}. \tag{4.18}$$

In order to prove Equation (4.18) we proceed by induction.

Assume that Equation (4.18) is true for a fixed $k = 1, \ldots, n-1$, and for all blocks $\mathbf{B}_{(0,j)}$ where $j < k$. Next we are going to show that Equation (4.18) is also valid for $k + 1$.

$$\begin{aligned}
\boldsymbol{\alpha}_{k+1:n} e^{\mathbf{G}_{k+1:n} y} &= (\overline{\boldsymbol{\alpha}}_n, \mathbf{0}) \begin{pmatrix} e^{\mathbf{G}_{k:n} y} & \displaystyle\int_0^y e^{\mathbf{G}_{k:n}(y-t)} \begin{pmatrix} \mathbf{0} \\ \mathbf{S}_{(k-1,k)} \end{pmatrix} e^{\mathbf{S}_{(k,k)} t} dt \\ \mathbf{0} & e^{\mathbf{S}_{(k,k)} y} \end{pmatrix} \\
&= \left( \boldsymbol{\alpha}_{k:n} e^{\mathbf{G}_{k:n} y}, \int_0^y \boldsymbol{\alpha}_{k:n} e^{\mathbf{G}_{k:n}(y-t)} \begin{pmatrix} \mathbf{0} \\ \mathbf{S}_{(k-1,k)} \end{pmatrix} e^{\mathbf{S}_{(k,k)} t} dt \right) \\
&= \begin{pmatrix} \boldsymbol{\alpha}_{k:n} e^{\mathbf{G}_{k:n} y}, \mathbf{B}_{(0,k)} \end{pmatrix}.
\end{aligned}$$

Now we are going to calculate the block $\mathbf{B}_{(0,k)}$ which is equal to the integral

$$\int_0^y \boldsymbol{\alpha}_{k:n} e^{\mathbf{G}_{k:n}(y-t)} \begin{pmatrix} \mathbf{0} \\ \mathbf{S}_{(k-1,k)} \end{pmatrix} e^{\mathbf{S}_{(k,k)} t} dt.$$

Consider the product

$$\begin{aligned}
& \boldsymbol{\alpha}_{k:n} e^{\mathbf{G}_{k:n}(y-t)} \begin{pmatrix} \mathbf{0} \\ \mathbf{S}_{(k-1,k)} \end{pmatrix} \\
= & \begin{pmatrix} \mathbf{B}_{(0,0)} & \mathbf{B}_{(0,1)} & \cdots & \mathbf{B}_{(0,k-1)} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{S}_{(k-1,k)} \end{pmatrix} \\
= & \mathbf{B}_{(0,k-1)} \mathbf{S}_{(k-1,k)}
\end{aligned}$$

$$= \left( \left( \underset{u\in\boldsymbol{\ell}}{\otimes} \boldsymbol{\alpha}_u e^{\mathbf{S}_u(y-t)} \right) \left( \prod_{v\in\boldsymbol{\ell}^c} F_v(y-t) \right) \right)_{\boldsymbol{\ell}\in\mathcal{C}_{n-(k-1),n}} \mathbf{S}_{(k-1,k)} \quad \text{(induction hypothesis)}$$

$$= \left( \left( \boldsymbol{\alpha}_w e^{\mathbf{S}_w(y-t)} \mathbf{s}_w \underset{u\in\boldsymbol{\gamma}}{\otimes} \boldsymbol{\alpha}_u e^{\mathbf{S}_u(y-t)} \right) \left( \prod_{v\in\boldsymbol{\ell}^c} F_v(y-t) \right) \right)_{\substack{\boldsymbol{\ell}\in\mathcal{C}_{n-(k-1),n},\ \boldsymbol{\gamma}\in\mathcal{C}_{n-k,n}, \\ \boldsymbol{\gamma}\subset\boldsymbol{\ell},\ \boldsymbol{\ell}-\boldsymbol{\gamma}=w}}$$

Now consider the next product

$$\boldsymbol{\alpha}_{k:n} e^{\mathbf{G}_{k:n}(y-t)} \begin{pmatrix} \mathbf{0} \\ \mathbf{S}_{(k-1,k)} \end{pmatrix} e^{\mathbf{S}_{(k,k)}t}$$

$$= \left( \left( \boldsymbol{\alpha}_w e^{\mathbf{S}_w(y-t)} \mathbf{s}_w \underset{u\in\boldsymbol{\gamma}}{\otimes} \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \right) \left( \prod_{v\in\boldsymbol{\ell}^c} F_v(y-t) \right) \right)_{\substack{\boldsymbol{\ell}\in\mathcal{C}_{n-(k-1),n},\ \boldsymbol{\gamma}\in\mathcal{C}_{n-k,n}, \\ \boldsymbol{\gamma}\subset\boldsymbol{\ell},\ \boldsymbol{\ell}-\boldsymbol{\gamma}=w}}$$

$$= \left( \left( \underset{u\in\boldsymbol{\gamma}}{\otimes} \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \right) \left( \boldsymbol{\alpha}_w e^{\mathbf{S}_w(y-t)} \mathbf{s}_w \prod_{v\in\boldsymbol{\ell}^c} F_v(y-t) \right) \right)_{\substack{\boldsymbol{\ell}\in\mathcal{C}_{n-(k-1),n},\ \boldsymbol{\gamma}\in\mathcal{C}_{n-k,n}, \\ \boldsymbol{\gamma}\subset\boldsymbol{\ell},\ \boldsymbol{\ell}-\boldsymbol{\gamma}=w}}$$

where

$$\begin{aligned} & \int_0^y \boldsymbol{\alpha}_w e^{\mathbf{S}_w(y-t)} \mathbf{s}_w \prod_{v\in\boldsymbol{\ell}^c} F_v(y-t)dt \\ = & \int_0^y f_w(y-t) \prod_{v\in\boldsymbol{\ell}^c} F_v(y-t)dt \\ = & \prod_{v\in\boldsymbol{\gamma}^c} F_v(y), \end{aligned}$$

where the last equality is due to the rule of differentiation of the product of functions.

Therefore, we finally obtain

$$\begin{aligned} \mathbf{B}_{(0,k)} &= \int_0^y \boldsymbol{\alpha}_{k:n} e^{\mathbf{G}_{k:n}(y-t)} \begin{pmatrix} \mathbf{0} \\ \mathbf{S}_{(k-1,k)} \end{pmatrix} e^{\mathbf{S}_{(k,k)}t} dt \\ &= \left( \left( \underset{u\in\boldsymbol{\gamma}}{\otimes} \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \right) \left( \prod_{v\in\boldsymbol{\gamma}^c} F_v(y) \right) \right)_{\boldsymbol{\gamma}\in\mathcal{C}_{n-k,n}}. \end{aligned}$$

$\square$

We finalize this section with the next Theorem where we make the last product of the calculations to prove that the proposed representations for order statistics given in Theorem 4.1 can also be applied to order statistics of ME-distributions.

**Theorem 4.4** *Let $Y_1, \ldots, Y_n$ be $n$ independent and ME-distributed random variables with $Y_j \sim ME(\boldsymbol{\alpha}_j, \mathbf{S}_j)$, $j = 1, \ldots, n$. Then, the $k$-th order statistic $Y_{(k:n)}$ $(k = 1, 2, \ldots, n)$ has a ME-representation given by $(\boldsymbol{\alpha}_{k:n}, \mathbf{G}_{k:n})$, where the row vector $\boldsymbol{\alpha}_{k:n}$ is given in Equation* (4.7) *and the matrix $\mathbf{G}_{k:n}$ is given in Equation* (4.8).

**Proof.** Let $y > 0$. Then

$$
\begin{aligned}
\boldsymbol{\alpha}_{k:n} e^{\mathbf{G}_{k:n} y} \mathbf{e} &= \begin{pmatrix} \mathbf{B}_{(0,0)} & \mathbf{B}_{(0,1)} & \cdots & \mathbf{B}_{(0,k-1)} \end{pmatrix} \mathbf{e} \\
&= \sum_{j=0}^{k-1} \mathbf{B}_{(0,j)} \mathbf{e} \\
&= \sum_{j=0}^{k-1} \left( \left( \bigotimes_{u \in \boldsymbol{\ell}} \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \right) \left( \prod_{v \in \boldsymbol{\ell}^c} F_v(y) \right) \right)_{\boldsymbol{\ell} \in \mathcal{C}_{n-j,n}} \mathbf{e}, \quad \text{by Lemma 4.3,} \\
&= \sum_{j=0}^{k-1} \left( \left( \prod_{u \in \boldsymbol{\ell}} \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \mathbf{e} \right) \left( \prod_{v \in \boldsymbol{\ell}^c} F_v(y) \right) \right)_{\boldsymbol{\ell} \in \mathcal{C}_{n-j,n}} \\
&= \sum_{j=0}^{k-1} \left( \left( \prod_{u \in \boldsymbol{\ell}} \overline{F}_u(y) \right) \left( \prod_{v \in \boldsymbol{\ell}^c} F_v(y) \right) \right)_{\boldsymbol{\ell} \in \mathcal{C}_{n-j,n}} \quad \text{where } \overline{F}_u(y) = \boldsymbol{\alpha}_u e^{\mathbf{S}_u y} \mathbf{e}, \\
&= \sum_{j=0}^{k-1} \frac{1}{j!(n-j)!} \operatorname{per}\big[ \underbrace{\mathbf{e} - \mathbf{F}(y)}_{\substack{n-j \\ \text{columns}}} \quad \underbrace{\mathbf{F}(y)}_{\substack{j \\ \text{columns}}} \big] \\
&= \mathbb{P}\left( Y_{(k:n)} > y \right) \quad \text{(see Equation (4.2)).}
\end{aligned}
$$

□

The proof in this section is originally given in [BN17, Section 4.4.2]. We have included it here to present a complete overview of order statistics from matrix-exponential distributions.

## 4.4 Order statistics from discrete Phase-type distributions

In this section we present two types of representations for order statistics from independent and DPH-distributed random variables. The first representation provided here corresponds to the discrete version of the representation given in Section 4.2 and the second representation results by considering a reduction of the dimension of the first representations for the case of i.i.d. random variables.

Representations for order statistics from DPH-distributions are certainly more tedious than the continuous PH-distributions. However, here we present a study of their representations by giving probabilistic interpretations. Also, all the notation introduced in this section will help for the following sections and analysis for the case of MG-distributions.

As an introduction to representations for order statistics from DPH-distributions we present a DPH-representation for the minimum and another one for the maximum of two independent and DPH-distributed random variables.

Let $Y_1$ and $Y_2$ be two independent and DPH-distributed random variables with representations given by $(\boldsymbol{\pi}_1, \mathbf{T}_1)$ and $(\boldsymbol{\pi}_2, \mathbf{T}_2)$ , respectively. Then, a DPH-representation for the minimum is given by

$$\left(\boldsymbol{\pi}_1 \otimes \boldsymbol{\pi}_2, \mathbf{T}_1 \otimes \mathbf{T}_2\right),$$

whereas a DPH-representation for the maximum is given by the pair

$$\left(\left(\boldsymbol{\pi}_1 \otimes \boldsymbol{\pi}_2, \mathbf{0}\right), \mathbf{P}_{(2)}\right),$$

where

$$\mathbf{P}_{(2)} = \begin{pmatrix} \mathbf{T}_1 \otimes \mathbf{T}_2 & \mathbf{T}_1 \otimes \mathbf{t}_2 & \mathbf{t}_1 \otimes \mathbf{T}_2 \\ \mathbf{0} & \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_2 \end{pmatrix}, \quad \mathbf{t}_i = \mathbf{e} - \mathbf{T}_i\mathbf{e}, \quad i = 1, 2,$$

and $\mathbf{0}$ denotes zero-matrices of appropriate dimension.

We can provide a probabilistic interpretation for the representation of the minimum and the maximum. Consider the multivariate Markov chain formed with the underlying Markov chains of $Y_1$ and $Y_2$. Then, the given representation for the minimum shows the initial probabilities and the transition probabilities when the two underlying Markov chains are in a transient state. The first time when only one of the underlying Markov chains gets absorbed or the two Markov chains get absorbed (at the same time) will correspond to the first order statistic. Therefore, an underlying Markov chain for the first order statistic has an initial distribution $\boldsymbol{\pi}_1 \otimes \boldsymbol{\pi}_2$ and sub-transition probability matrix $\mathbf{T}_1 \otimes \mathbf{T}_2$.

Consider now the representation given for the maximum. This representation considers the transition probabilities of the representation of the first order statistic and it also takes into account the transition probabilities between states of the multivariate Markov chain which expresses that one of the underlying Markov chains has been absorbed but the other one is still running in its transient states. Thus, the initial distribution of the multivariate Markov chain has an initial distribution $(\boldsymbol{\pi}_1 \otimes \boldsymbol{\pi}_2, \mathbf{0})$ and sub-transition probability matrix $\mathbf{P}_{(2)}$.

Under that framework, order statistics have an inherent probabilistic interpretation that leads to construct representations for order statistics from a finite set of independent and DPH-distributed random variables. In the following, we present the DPH-representations for general order statistics.

Let $\left\{X_t^1\right\}_{t\in\mathbb{N}}, \left\{X_t^2\right\}_{t\in\mathbb{N}}, \ldots, \{X_t^n\}_{t\in\mathbb{N}}$ be $n$ independent Markov chains with state space $\mathcal{E}_j, j = 1, \ldots, n$, respectively. Every $\mathcal{E}_j$ is given by $\left\{1, \ldots, p_j, p_j + 1\right\}$, where $\left\{1, \ldots, p_j\right\}$ are transient states and $p_j + 1$ is the absorbing state of the corresponding Markov chain $\left\{X_t^j\right\}_{t\in\mathbb{N}}$.

For each $j = 1, \ldots, n$, let $\boldsymbol{\pi}_j$ be the initial distribution and let

$$\mathbf{P}_j = \begin{pmatrix} \mathbf{T}_j & \mathbf{t}_j \\ \mathbf{0} & 1 \end{pmatrix}, \quad \mathbf{t}_j = \mathbf{e} - \mathbf{T}_j\mathbf{e},$$

be the transition probability matrix of $\left\{X_t^j\right\}_{t\in\mathbb{N}}$.

Consider $\mathbf{X}_t = \left(X_t^1, X_t^2, \ldots, X_t^n\right), t \in \mathbb{N}$, (this is $n$ independent Markov chains running in parallel), where its state space is denoted by

$$\mathcal{E} = \bigotimes_{j=1}^{n} \mathcal{E}_j.$$

The Kronecker product $\otimes$ provides a lexicographical ordering for the state space $\mathcal{E}$; nevertheless, we are going to consider another ordering for $\mathcal{E}$, which is going to be more convenient and it consists on making a classification of the multi-dimensional states and later ordering the states lexicographically. For the classification we group the multi-dimensional states according to the number of absorbing states in their entries, as it is shown next.

For each $j = 0, 1, \ldots, n$, let

$$\mathcal{A}_j = \left\{ (\xi_1, \ldots, \xi_n) \in \mathcal{E} \,\middle|\, \sum_{l=1}^{n} \mathbf{1}_{(\xi_l)} = j \right\}, \quad \text{where} \quad \mathbf{1}_{(\xi_l)} = \begin{cases} 1 & \text{if} \quad \xi_l = p_l + 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that all the subsets $\left\{\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_n\right\}$ together form a partition of the state space $\mathcal{E}$.

Since the multivariate Markov chain $\mathbf{X}_t$ starts on the subset $\mathcal{A}_0$, then all of the initial probabilities are enclosed in $(\boldsymbol{\pi}_1 \otimes \cdots \otimes \boldsymbol{\pi}_n, \mathbf{0})$, where $\mathbf{0}$ is a vector of zeros of dimension $\sum_{j=1}^{n} \dim(\mathcal{A}_j)$.

Let

$$\mathbf{P} = \begin{pmatrix} \mathbf{C}_{(0,0)} & \mathbf{C}_{(0,1)} & \mathbf{C}_{(0,2)} & \cdots & \mathbf{C}_{(0,n-2)} & \mathbf{C}_{(0,n-1)} & \mathbf{r}_0 \\ \mathbf{0} & \mathbf{C}_{(1,1)} & \mathbf{C}_{(1,2)} & \cdots & \mathbf{C}_{(1,n-2)} & \mathbf{C}_{(1,n-1)} & \mathbf{r}_1 \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{(2,2)} & \cdots & \mathbf{C}_{(2,n-2)} & \mathbf{C}_{(2,n-1)} & \mathbf{r}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{(n-2,n-2)} & \mathbf{C}_{(n-2,n-1)} & \mathbf{r}_{n-2} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{(n-1,n-1)} & \mathbf{r}_{n-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \tag{4.19}$$

be the transition probability matrix of $\mathbf{X}_t$. Below we explain the matrix $\mathbf{P}$.

The first sub-matrix is given by

$$\mathbf{C}_{(0,0)} = \mathbf{T}_1 \otimes \cdots \otimes \mathbf{T}_n. \tag{4.20}$$

Notice that $\mathbf{C}_{(0,0)}$ contains all the transition probabilities among the multi-dimensional states of the set $\mathcal{A}_0$.

In order to explain the rest of the sub-matrices in $\mathbf{P}$ we introduce again the set of combinations written in a vector form. For a fixed $i = 1, 2, \ldots, n-1$, consider the set of all combinations with $i$ elements of $\{1, 2, \ldots, n\}$, where we write every combination in a vector form as follows

$$\mathcal{C}_{i,n} = \left\{ \left(c_1^\theta, c_2^\theta, \ldots, c_i^\theta\right) : c_i^\theta \in \{1, 2, \ldots, n\}, \quad \theta \in \{1, \ldots, \tfrac{n!}{i!(n-i)!}\} \right\},$$

and every element in $\mathcal{C}_{i,n}$ is called $i$-combination. We set a lexicographical ordering on the set $\mathcal{C}_{i,n}$.

Notice that every $i$-combination in $\mathcal{C}_{i,n}$ is marked with a superscript which helps to name an $i$-combination. For example, for the $i$-combination $\left(c_1^\ell, c_2^\ell, \ldots, c_i^\ell\right)$ we refer to it as the $\boldsymbol{\ell}$-tuple. Also, for the $i$-combination $\left(c_1^\ell, c_2^\ell, \ldots, c_i^\ell\right)$ we say that $c_k^\ell \in \boldsymbol{\ell}$ for all $k = 1, \ldots, i$.

Every block $\mathbf{C}_{(0,i)}$ $(i = 1, 2, \ldots, n-1,)$ is defined as a row block matrix given by

$$\mathbf{C}_{(0,i)} = \left\{ \bigotimes_{\jmath=1}^{n} g_\jmath^{\mathbf{1}} \quad \bigotimes_{\jmath=1}^{n} g_\jmath^{\mathbf{2}} \quad \cdots \quad \bigotimes_{\jmath=1}^{n} g_\jmath^{\mathbf{C_i^n}} \right\} \tag{4.21}$$

$$\text{and for every } \boldsymbol{\gamma} \in \mathcal{C}_{i,n} \quad g_\jmath^{\boldsymbol{\gamma}} = \begin{cases} \mathbf{t}_\jmath & \text{if } \jmath \in \boldsymbol{\gamma} \\ \mathbf{T}_\jmath & \text{if } \jmath \notin \boldsymbol{\gamma} \end{cases}.$$

Thus, $\mathbf{C}_{(0,i)}$ contains $\binom{n}{i}$ Kronecker products placed consecutively.

Every block $\mathbf{C}_{(i,i)}$ $(i = 1, 2, \ldots, n-1, )$ denotes a diagonal block matrix formed by

$$\mathbf{C}_{(i,i)} = \text{diag}\left\{ \overset{n}{\underset{\jmath=1}{\otimes}} g_\jmath^{\gamma} : \quad \gamma \in \mathcal{C}_{i,n} \right\}, \quad g_\jmath^{\gamma} = \left\{ \begin{array}{ll} 1 & \text{if } \jmath \in \boldsymbol{\gamma} \\ \mathbf{T}_\jmath & \text{if } \jmath \notin \boldsymbol{\gamma} \end{array} \right. . \tag{4.22}$$

Here there are $\binom{n}{i}$ Kronecker products located in the diagonal in alignment with the corresponding Kronecker product in $\mathbf{C}_{(0,i)}$ (with respect to the lexicographical ordering).

For every fixed $i = 1, \ldots, n-2$, and $s = i+1, \ldots, n-1$, then $\mathbf{C}_{(i,s)}$ is formed by

$$\left[\mathbf{C}_{(i,s)}\right]_{\boldsymbol{\gamma},\boldsymbol{\ell}} = \left\{ \overset{n}{\underset{\jmath=1}{\otimes}} g_\jmath^{\boldsymbol{\gamma},\boldsymbol{\ell}} : \quad \gamma \in \mathcal{C}_{i,n}, \quad \boldsymbol{\ell} \in \mathcal{C}_{s,n} \right\}, \quad g_\jmath^{\boldsymbol{\gamma},\boldsymbol{\ell}} = \left\{ \begin{array}{ll} 1 & \text{if } \jmath \in \boldsymbol{\gamma}, \jmath \in \boldsymbol{\ell} \\ \mathbf{t}_\jmath & \text{if } \jmath \notin \boldsymbol{\gamma}, \jmath \in \boldsymbol{\ell} \\ \mathbf{T}_\jmath & \text{if } \jmath \notin \boldsymbol{\gamma}, \jmath \notin \boldsymbol{\ell} \\ \mathbf{0} & \text{if } \jmath \in \boldsymbol{\gamma}, \jmath \notin \boldsymbol{\ell} \end{array} \right. , \tag{4.23}$$

where every block entry is placed depending on the location of the chosen vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\ell}$. Thus, in $\mathbf{C}_{(i,s)}$ there are $\binom{n}{i} \times \binom{n}{s}$ block entries in total.

Finally, since $\mathbf{P}$ is a block-partitioned transition probability matrix, then we have

$$\mathbf{r}_i = \mathbf{e} - \sum_{l=i}^{n-1} \mathbf{C}_{(i,l)} \mathbf{e}, \quad i = 0, 1, \ldots, n-1.$$

We will need to introduce some notation to describe the power of the matrix $\mathbf{P}$.

Let

$$\mathbf{P}^m = \begin{pmatrix} \mathbf{C}_{(0,0,m)} & \mathbf{C}_{(0,1,m)} & \mathbf{C}_{(0,2,m)} & \cdots & \mathbf{C}_{(0,n-2,m)} & \mathbf{C}_{(0,n-1,m)} & \mathbf{r}_{0,m} \\ \mathbf{0} & \mathbf{C}_{(1,1,m)} & \mathbf{C}_{(1,2,m)} & \cdots & \mathbf{C}_{(1,n-2,m)} & \mathbf{C}_{(1,n-1,m)} & \mathbf{r}_{1,m} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{(2,2,m)} & \cdots & \mathbf{C}_{(2,n-2,m)} & \mathbf{C}_{(2,n-1,m)} & \mathbf{r}_{2,m} \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{(n-2,n-2,m)} & \mathbf{C}_{(n-2,n-1,m)} & \mathbf{r}_{n-2,m} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{(n-1,n-1,m)} & \mathbf{r}_{n-1,m} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & 1 \end{pmatrix}$$

denote the $m$-th power of $\mathbf{P}$.

The first sub-matrix in the diagonal of $\mathbf{P}^m$ is given by

$$\mathbf{C}_{(0,0,m)} = \mathbf{C}_{(0,0)}^m = \mathbf{T}_1^m \otimes \cdots \otimes \mathbf{T}_n^m. \tag{4.24}$$

The rest of the matrices in the diagonal, $\left\{\mathbf{C}_{(i,i,m)}, i = 1, \ldots, n-1\right\}$, are diagonal matrices by themselves and every entry corresponds to one block matrix formed by

Kronecker products. Thus, the power of every entry is calculated by applying $(m-1)$-times the Kronecker product property (see Equation (A.6)).

The matrices $\left\{ \mathbf{C}_{(0,i,m)}, i = 1, \dots, n-1, \right\}$ can be calculated by iteration as follows

$$\mathbf{C}_{(0,i,m)} = \sum_{k=0}^{m-1} \sum_{l=0}^{i-1} \mathbf{C}_{(0,l,k)} \mathbf{C}_{(l,i)} \mathbf{C}_{(i,i,m-1-k)}. \tag{4.25}$$

For every fixed $i = 1, \dots, n-1$ the matrices $\left\{ \mathbf{C}_{(i,s,m)}, s = i+1, \dots, n-1, \right\}$ can also be calculated by iteration as follows

$$\mathbf{C}_{(i,s,m)} = \sum_{k=0}^{m-1} \sum_{l=i}^{s-1} \mathbf{C}_{(i,l,k)} \mathbf{C}_{(l,s)} \mathbf{C}_{(s,s,m-1-k)}. \tag{4.26}$$

Finally, since $\mathbf{P}^m$ is a block-partitioned transition probability matrix, then we have

$$\mathbf{r}_{i,m} = \mathbf{e} - \sum_{l=i}^{n-1} \mathbf{C}_{(i,l,m)} \mathbf{e}, \quad i = 0, \dots, n-1.$$

Let $Y_1, Y_2, \dots, Y_n$ be independent and DPH-distributed random variables with representation given by $(\boldsymbol{\pi}_i, \mathbf{T}_i)$ , $i = 1, \dots, n$, respectively.

From the probabilistic interpretation of the matrix $\mathbf{P}^m$, we have

$$\begin{aligned} \mathbb{P}\left(Y_{(1:n)} > m\right) &= \mathbb{P}\left(\mathbf{X}_m \in \mathcal{A}_0\right) = \overline{\boldsymbol{\pi}}_n \mathbf{C}_{(0,0,m)} \mathbf{e}, \\ \mathbb{P}\left(Y_{(i-1:n)} \leq m, Y_{(i:n)} > m\right) &= \mathbb{P}\left(\mathbf{X}_m \in \mathcal{A}_{i-1}\right) = \overline{\boldsymbol{\pi}}_n \mathbf{C}_{(0,i-1,m)} \mathbf{e}, \end{aligned}$$

where $\overline{\boldsymbol{\pi}}_n = \boldsymbol{\pi}_1 \otimes \cdots \otimes \boldsymbol{\pi}_n$ and $i = 2, \dots, n-1$.

Then, for a fixed $r \in \{1, 2, \dots, n\}$ we obtain

$$\mathbb{P}\left(Y_{(r:n)} > m\right) = \overline{\boldsymbol{\pi}}_n \mathbf{C}_{(0,0,m)} \mathbf{e} + \overline{\boldsymbol{\pi}}_n \mathbf{C}_{(0,1,m)} \mathbf{e} + \cdots + \overline{\boldsymbol{\pi}}_n \mathbf{C}_{(0,r-1,m)} \mathbf{e}. \tag{4.27}$$

Hence, the survival function of $Y_{(r:n)}$ depends only on the $r$ block matrices located in the first row block of $\mathbf{P}$ (see Equation (4.19)). This means that by taking the matrix

$$\mathbf{P}_{(r)} = \begin{pmatrix} \mathbf{C}_{(0,0)} & \mathbf{C}_{(0,1)} & \mathbf{C}_{(0,2)} & \cdots & \mathbf{C}_{(0,r-2)} & \mathbf{C}_{(0,r-1)} \\ \mathbf{0} & \mathbf{C}_{(1,1)} & \mathbf{C}_{(1,2)} & \cdots & \mathbf{C}_{(1,r-2)} & \mathbf{C}_{(1,r-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{(2,2)} & \cdots & \mathbf{C}_{(2,r-2)} & \mathbf{C}_{(2,r-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{(r-2,r-2)} & \mathbf{C}_{(r-2,r-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{(r-1,r-1)} \end{pmatrix} \tag{4.28}$$

the expression in Equation (4.27) can be written in the following form

$$\mathbb{P}\left(Y_{(r:n)} > m\right) = (\overline{\boldsymbol{\pi}}_n, \mathbf{0})\mathbf{P}_{(r)}^m \mathbf{e}, \tag{4.29}$$

where $\mathbf{0}$ is a vector of zeros with an appropriate dimension with respect to the matrix $\mathbf{P}_{(r)}$. Therefore, we conclude that one DPH-representation for $Y_{(r:n)}$ is $\left((\overline{\boldsymbol{\pi}}_n, \mathbf{0}), \mathbf{P}_{(r)}\right)$.

### 4.4.1 Order reduction for DPH-representations

Here, we present an order reduction for the DPH-representations of order statistics when the random variables are identically distributed.

Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically discrete phase-type distributed with representation $\text{DPH}_p(\boldsymbol{\pi}, \mathbf{T})$. We recall the multivariate Markov chain

$$\mathbf{X}_t = \left(X_t^1, X_t^2, \ldots, X_t^n\right), t \in \mathbb{N},$$

where $X_t^i$ is the corresponding underlying Markov chain of $Y_i$ and $\mathcal{E}$ denotes its state space.

For every $s = 1, \ldots, p$, let $\mathbf{m}_s$ be a function defined as follows:

$$\mathbf{m}_s(\xi_1, \ldots, \xi_n) = \sum_{j=1}^{n} \mathbf{1}_{\{\xi_j = s\}}, \quad (\xi_1, \ldots, \xi_n) \in \mathcal{E},$$

thus, $\mathbf{m}_s$ counts the number of Markov chains $X_t^i$ in state $s$ when the multivariate Markov chain $\mathbf{X}_t$ is in the state $(\xi_1, \ldots, \xi_n)$.

Now, let us define the vector given by

$$\overline{\mathbf{m}}(\xi_1, \ldots, \xi_n) = (\mathbf{m}_1(\xi_1, \ldots, \xi_n), \ldots, \mathbf{m}_p(\xi_1, \ldots, \xi_n)),$$

which is a mapping from the state space $\mathcal{E}$ to a set of p-dimensional states where what only matters is the number of Markov chains that are still running (or the ones that have not become absorbed).

For every $s = 1, \ldots, p$, let $N_s(m)$ denote the number of Markov chains $X_i's$ in the state $s$ at time $m$. Then

$$\mathbf{N}(m) = (N_1(m), N_2(m), \ldots, N_p(m)), \quad m \in \mathbb{N}$$

is a multivariate Markov chain. We denote by $\mathcal{M}$ its state space and by $\mathbf{Q}$ its transition probability matrix.

For every $\mathbf{n} = (n_1, n_2, \ldots, n_p) \in \mathcal{M}$ we define

$$\mathcal{E}_{\mathbf{n}} = \{ (\xi_1, \xi_2, \ldots, \xi_n) \in \mathcal{E} | \quad \overline{\mathbf{m}} (\xi_1, \xi_2, \ldots, \xi_n) = \mathbf{n} \} .$$

Notice that $\mathcal{E}_{\mathbf{n}}$ consists on permutations of the set $\{\xi_1, \xi_2, \ldots, \xi_n\}$ with cardinality $\frac{n!}{n_1! n_2! \cdots n_p!}$.

Now, for every pair $\mathbf{n}_1, \mathbf{n}_2 \in \mathcal{M}$ we have that

$$\begin{aligned} & \mathbb{P} (\mathbf{N}(1) = \mathbf{n}_2 | \mathbf{N}(0) = \mathbf{n}_1 ) \\ = & \frac{\mathbb{P} (\mathbf{N}(1) = \mathbf{n}_2, \mathbf{N}(0) = \mathbf{n}_1 )}{\mathbb{P} (\mathbf{N}(0) = \mathbf{n}_1 )} \\ = & \frac{\sum\limits_{\mathbf{u} \in \mathcal{E}_{\mathbf{n}_1}} \sum\limits_{\mathbf{v} \in \mathcal{E}_{\mathbf{n}_2}} \mathbb{P} \left( \mathbf{X}_1 = \mathbf{v} \middle| \mathbf{X}_0 = \mathbf{u} \right) \mathbb{P} (\mathbf{X}_0 = \mathbf{u})}{\sum\limits_{\mathbf{k} \in \mathcal{E}_{\mathbf{n}_1}} \mathbb{P} (\mathbf{X}_0 = \mathbf{k})} . \end{aligned} \tag{4.30}$$

$\mathbb{P} (\mathbf{X}_0 = \mathbf{k}_1) = \mathbb{P} (\mathbf{X}_0 = \mathbf{k}_2)$, for all $\mathbf{k}_1, \mathbf{k}_2 \in \mathcal{E}_{\mathbf{n}_1}$, since $\mathbf{k}_1$ is a permutation of the entries of $\mathbf{k}_2$ and $Y_1, \ldots, Y_n$ are independent and identically distributed. Then by letting $u$ be the cardinality of the set $\mathcal{E}_{\mathbf{n}_1}$, we get that

$$\frac{\mathbb{P} (\mathbf{X}_0 = \mathbf{k}_1)}{\sum\limits_{\mathbf{k} \in \mathcal{E}_{\mathbf{n}_1}} \mathbb{P} (\mathbf{X}_0 = \mathbf{k})} = \frac{1}{u},$$

and consequently Equation $(4.30)$ becomes to

$$\mathbb{P} (\mathbf{N}(1) = \mathbf{n}_2 | \mathbf{N}(0) = \mathbf{n}_1 ) = \frac{1}{u} \sum_{\mathbf{u} \in \mathcal{E}_{\mathbf{n}_1}} \sum_{\mathbf{v} \in \mathcal{E}_{\mathbf{n}_2}} \mathbb{P} \left( \mathbf{X}_1 = \mathbf{v} \middle| \mathbf{X}_0 = \mathbf{u} \right) . \tag{4.31}$$

Let $\mathbf{u} = (\xi_1, \ldots, \xi_n) \in \mathcal{E}_{\mathbf{n}_1}$ and $\mathbf{v} = (h_1, \ldots, h_n) \in \mathcal{E}_{\mathbf{n}_2}$. Then

$$\mathbb{P} \left( \mathbf{X}_1 = \mathbf{v} \middle| \mathbf{X}_0 = \mathbf{u} \right) = t_{\xi_1, h_1} t_{\xi_2, h_2} \cdots t_{\xi_n, h_n} .$$

From here, observe that if we consider a permutation of the elements $\{\xi_1, \ldots, \xi_n\}$, let say $(\xi_{\rho_1}, \ldots, \xi_{\rho_n})$, then we can find another permutation of the elements $\{h_1, \ldots, h_n\}$, let say $(h_{\rho_1}, \ldots, h_{\rho_n})$, such that

$$t_{\xi_1, h_1} t_{\xi_2, h_2} \cdots t_{\xi_n, h_n} = t_{\xi_{\rho_1}, h_{\rho_1}} t_{\xi_{\rho_2}, h_{\rho_2}} \cdots t_{\xi_{\rho_n}, h_{\rho_n}} .$$

Thus, there exits state $\widehat{\mathbf{u}} = (\xi_{\rho_1}, \ldots, \xi_{\rho_n}) \in \mathcal{E}_{\mathbf{n}_1}$ and state $\widehat{\mathbf{v}} = (h_{\rho_1}, \ldots, h_{\rho_n}) \in \mathcal{E}_{\mathbf{n}_2}$ such that

$$\mathbb{P} \left( \mathbf{X}_1 = \mathbf{v} \middle| \mathbf{X}_0 = \mathbf{u} \right) = \mathbb{P} \left( \mathbf{X}_1 = \widehat{\mathbf{v}} \middle| \mathbf{X}_0 = \widehat{\mathbf{u}} \right) .$$

Therefore, Equation (4.31) now becomes

$$\mathbb{P}\left(\mathbf{N}(1)=\mathbf{n}_2 \left|\mathbf{N}(0)=\mathbf{n}_1\right.\right)=\sum_{\mathbf{v}\in\mathcal{E}_{\mathbf{n}_2}} \mathbb{P}\left(\mathbf{X}_1=\mathbf{v}\middle|\mathbf{X}_0=\mathbf{u}\right), \tag{4.32}$$

where $\mathbf{u}$ is any state in $\mathcal{E}_{\mathbf{n}_1}$.

Equation (4.32) shows one relation between the transition matrix $\mathbf{Q}$ and the transition matrix $\mathbf{P}$. In the next, we are going to see into more details the transformation between the two transition probability matrices.

It is convenient for us to define an ordering for the states in $\mathcal{M}$. We first order them decreasingly according to the number of Markov chains that are still running, this is according to the number $\widehat{\mathbf{m}}\left(\xi_1,\ldots,\xi_n\right)\mathbf{e}$. Secondly, we group all the states that have the same number $\widehat{\mathbf{m}}\left(\xi_1,\ldots,\xi_n\right)\mathbf{e}$ and later we consider a lexicographical ordering among them.

Let $\mathbf{M}=\{m_{i,j}\}$ be a matrix of dimension $l\times m$, where $l$ is equal to the number of states in $\mathcal{E}$ and $m$ is the number of states in $\mathcal{M}$. Let $\mathbf{u}_i$ be the i-th state of $\mathcal{E}$ and let $\mathbf{v}_j$ denote the j-th state of $\mathcal{M}$. Then,

$$m_{i,j}=\begin{cases} 1 & \text{if} \quad \overline{\mathbf{m}}\left(\mathbf{u}_i\right)=\mathbf{v}_i, \\ 0 & \text{otherwise.}\end{cases} \tag{4.33}$$

Observe that every row in the matrix $\mathbf{M}$ has only one entry with value of 1 and the rest of the entries are zero.

**Lemma 4.5** $\mathbf{PM}=\mathbf{MQ}$.

**Proof.** Let us consider again the states $\mathbf{u}_i\in\mathcal{E}$ and $\mathbf{v}_j\in\mathcal{M}$. Then, on the one hand we have

$$\left(\mathbf{PM}\right)_{i,j}=\sum_{k\in\mathcal{E}_{\mathbf{v}_j}}\mathbf{p}_{i,k}, \tag{4.34}$$

where the sum is over all the states that belong to $\mathcal{E}_{\mathbf{v}_j}$.

On the other hand, let us denote by $\mathbf{n}$ the state in $\mathcal{M}$ such that $\overline{\mathbf{m}}\left(\mathbf{u}_i\right)=\mathbf{n}$. Then,

$$\left(\mathbf{MQ}\right)_{i,j}=\mathbb{P}\left(\mathbf{N}(1)=\mathbf{v}_j\middle|\mathbf{N}(0)=\mathbf{n}\right), \tag{4.35}$$

and by using Equation (4.32) we get

$$\left(\mathbf{MQ}\right)_{i,j} \quad = \quad \sum_{\mathbf{v}\in\mathcal{E}_{\mathbf{v}_j}}\mathbb{P}\left(\mathbf{X}_1=\mathbf{v}\middle|\mathbf{X}_0=\mathbf{u}\right), \quad \text{where } \mathbf{u} \text{ is any state in } \mathcal{E}_{\mathbf{n}_1},$$

$$
\begin{aligned}
&= \sum_{\mathbf{v}\in\mathcal{E}_{\mathbf{v}_j}} \mathbb{P}\left(\mathbf{X}_1 = \mathbf{v}\middle|\mathbf{X}_0 = \mathbf{u}_i\right) \\
&= (\mathbf{PM})_{i,j} .
\end{aligned}
$$

□

Consider the matrix $\mathbf{P}_{(r)}$ defined in Equation (4.28) and we denote its dimension by $p_r \times p_r$. Also consider the matrix $\mathbf{M}$ defined in Equation (4.33) of dimension $m_1 \times m_2$. We define the matrix $\mathbf{M}_{(r)}$ as the sub-matrix of $\mathbf{M}$ which is located in the upper left corner of dimension $p_r \times m_2$.

**Corollary 4.6** *Let* $\left((\overline{\boldsymbol{\pi}}_n, \mathbf{0}), \mathbf{P}_{(r)}\right)$ *be the representation for $r$-th order statistics defined in Equation* (4.28). *Let* $\mathbf{Q}_{(r)}$ *be such that it satisfies* $\mathbf{P}_{(r)}\mathbf{M}_{(r)} = \mathbf{M}_{(r)}\mathbf{Q}_{(r)}$ *and* $\mathbf{q}_{(r)} = (\overline{\boldsymbol{\pi}}_n, \mathbf{0})\,\mathbf{M}_{(r)}$. *Then,* $\left(\mathbf{q}_{(r)}, \mathbf{Q}_{(r)}\right)$ *is another representation for the $r$-th order statistics.*

**Proof.** Let $m \in \mathbb{N}$. By using that $\mathbf{M}_{(r)}\mathbf{Q}_{(r)} = \mathbf{P}_{(r)}\mathbf{M}_{(r)}$, then inductively we obtain

$$
\begin{aligned}
\mathbf{q}_{(r)}\mathbf{Q}_{(r)}^m\mathbf{e} &= (\overline{\boldsymbol{\pi}}_n, \mathbf{0})\,\mathbf{M}_{(r)}\mathbf{Q}_{(r)}^m\mathbf{e} = (\overline{\boldsymbol{\pi}}_n, \mathbf{0})\,\mathbf{M}_{(r)}\mathbf{Q}_{(r)}\mathbf{Q}_{(r)}^{m-1}\mathbf{e} \\
&= (\overline{\boldsymbol{\pi}}_n, \mathbf{0})\,\mathbf{P}_{(r)}\mathbf{M}_{(r)}\mathbf{Q}_{(r)}^{m-1}\mathbf{e} \\
&= (\overline{\boldsymbol{\pi}}_n, \mathbf{0})\,\mathbf{P}_{(r)}^m\mathbf{M}_{(r)}\mathbf{e} \\
&= (\overline{\boldsymbol{\pi}}_n, \mathbf{0})\,\mathbf{P}_{(r)}^m\mathbf{e}
\end{aligned}
$$

where the last equality is due to $\mathbf{M}_{(r)}\mathbf{e} = \mathbf{e}$. □

## 4.5 Order statistics from Matrix-geometric distributions

In this section we present two types of representations for order statistics from matrix-geometric distributions and another one derived by considering an order reduction as in the case of DPH-distributions.

### 4.5.1 First type of MG-representations

The proof of the representations for order statistics from discrete phase-type distribution given in Section 4.4 helps us to explain that the same form of the representations can be applied to order statistics from matrix-geometric distributions.

Let $Y_1, Y_2, \ldots, Y_n$ be independent and MG-distributed random variables where $Y_j \sim \mathrm{MG}\left(\boldsymbol{\alpha}_j, \mathbf{S}_j, \mathbf{s}_j\right)$ and $\mathbf{s}_j = \mathbf{e} - \mathbf{Se}$ for all $j = 1, \ldots, n$. Let $f_j(m)$ and $F_j(m)$ denote the probability function and the distribution function, respectively, of $Y_j$, for all $j = 1, \ldots, n$.

Recall Equation (4.2) and consider one term of its sum, let say

$$\frac{1}{i!(n-i)!}\mathrm{per}\big[\underbrace{\mathbf{F}(m)}_{\substack{i\\ \text{columns}}} \quad \underbrace{\mathbf{e}-\mathbf{F}(m)}_{\substack{n-i\\ \text{columns}}}\big]. \tag{4.36}$$

Then, notice that one resulting product from the expression (4.36) is given by

$$F_1(m)F_2(m)\cdots F_i(m)\big(1-F_{i+1}(m)\big)\cdots\big(1-F_n(m)\big),$$

which is identified by the $i$-combination $(1, 2, \ldots, i)$ and it is equivalent to

$$\sum_{(c_1^l, c_2^l, \ldots, c_i^l) \in \mathcal{P}_{i,(1,m)}^{rep}} f_1(c_1^l) f_2(c_2^l) \cdots f_i(c_i^l)\big(1-F_{i+1}(m)\big)\cdots\big(1-F_n(m)\big) \tag{4.37}$$

where $\mathcal{P}_{i,(1,m)}^{rep}$ is the set of all permutations (with repetitions) of size $i$ of the set $\{1, 2, \ldots, n\}$. Now, Equation (4.37) can also be written as

$$\sum_{(c_1^l, c_2^l, \ldots, c_i^l) \in \mathcal{P}_{i,(0,m-1)}^{rep}} (\boldsymbol{\alpha}_1 \mathbf{S}_1^{c_1^l}\mathbf{s}_1)(\boldsymbol{\alpha}_2 \mathbf{S}_2^{c_2^l}\mathbf{s}_2)\cdots(\boldsymbol{\alpha}_i \mathbf{S}_i^{c_i^l}\mathbf{s}_i)(\boldsymbol{\alpha}_{i+1}\mathbf{S}_{i+1}^{m}\mathbf{e})\cdots(\boldsymbol{\alpha}_n \mathbf{S}_n^{m}\mathbf{e}), \tag{4.38}$$

and finally we can write Equation (4.38) as

$$\sum_{(c_1^l, c_2^l, \ldots, c_i^l) \in \mathcal{P}_{i,(0,m-1)}^{rep}} \overline{\boldsymbol{\alpha}}_n \big(\mathbf{S}_1^{c_1^l}\mathbf{s}_1 \otimes \mathbf{S}_2^{c_2^l}\mathbf{s}_2 \otimes \cdots \otimes \mathbf{S}_i^{c_i^l}\mathbf{s}_i \otimes \mathbf{S}_{i+1}^{m} \otimes \cdots \otimes \mathbf{S}_n^{m}\big)\mathbf{e}, \tag{4.39}$$

where $\overline{\boldsymbol{\alpha}}_n = \boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n$.

In the following, we define a row-block matrix $\mathbf{B}_{(0,i,m)}$ for every $i = 1, 2, \ldots, n-1$.

Observe that for every $\boldsymbol{\ell} \in \mathcal{C}_{i,n}$ we get a set of vectors $c_{\boldsymbol{\ell}} = \left(c_1^{\ell}, c_2^{\ell}, \ldots, c_i^{\ell}\right)$ which refer to set of permutations (with repetitions) of size $i$ from the set $\{0, 1, \ldots, m-1\}$. Thus, we construct the row-block matrix

$$\mathbf{B}_{(0,i,m)} = \left(\sum_{C_{\mathbf{1}} \in \mathcal{P}_{i,(0,m-1)}^{rep}} \bigotimes_{\jmath=1}^{n} g_{\jmath}^{\mathbf{1}} \quad \sum_{C_{\mathbf{2}} \in \mathcal{P}_{i,(0,m-1)}^{rep}} \bigotimes_{\jmath=1}^{n} g_{\jmath}^{\mathbf{2}} \quad \cdots \quad \sum_{C_{\mathbf{C_i^n}} \in \mathcal{P}_{i,(0,m-1)}^{rep}} \bigotimes_{\jmath=1}^{n} g_{\jmath}^{\mathbf{C_i^n}}\right), \tag{4.40}$$

where for every $\boldsymbol{\ell} \in \mathcal{C}_{i,n}$ we define

$$g_{\jmath}^{\boldsymbol{\ell}} = \begin{cases} \mathbf{S}_{\jmath}^{c_{\jmath}^{\ell}}\mathbf{s}_{\jmath} & \text{if } \jmath \in \boldsymbol{\ell} \\ \mathbf{S}_{\jmath}^{m} & \text{if } \jmath \notin \boldsymbol{\ell} \end{cases},$$

and where $\sum\limits_{C_{\boldsymbol{\ell}} \in \mathcal{P}^{rep}_{i,(0,m-1)}}$ is a sum over all the values of the powers $c^{\ell}_{\jmath}$ of the matrices $\mathbf{S}_{\jmath}, \jmath \in \boldsymbol{\ell}$, (this is $m^i$ Kronecker products in the sum).

In this way, by multiplying the row-block matrix in Equation (4.40) with $\overline{\boldsymbol{\alpha}}_n = \boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n$ and then with $\mathbf{e}$, we obtain

$$\overline{\boldsymbol{\alpha}}_n \mathbf{B}_{(0,i,m)} \mathbf{e} = \frac{1}{i!(n-i)!} \mathrm{per}\big[\underbrace{\mathbf{F}(m)}_{\substack{i \\ \text{columns}}} \quad \underbrace{\mathbf{e} - \mathbf{F}(m)}_{\substack{n-i \\ \text{columns}}}\big].$$

Consequently

$$\mathbb{P}\left(Y_{(r:n)} > m\right) = \sum_{i=0}^{r-1} \overline{\boldsymbol{\alpha}}_n \mathbf{B}_{(0,i,m)} \mathbf{e} = \overline{\boldsymbol{\alpha}}_n \left(\mathbf{B}_{(0,0,m)}, \mathbf{B}_{(0,1,m)}, \ldots, \mathbf{B}_{(0,r-1,m)}\right)\mathbf{e},$$

see Equation (4.2).

Now, in order to get a representation for the $r$-th order statistic $Y_{(r:n)}$, we look for one matrix $\mathbf{A}$ and one vector $\mathbf{v}$ such that

$$\mathbf{v}\mathbf{A}^m \mathbf{e} = \sum_{i=0}^{r-1} \overline{\boldsymbol{\alpha}}_n \mathbf{B}_{(0,i,m)} \mathbf{e}.$$

Naturally the first suggestion is the matrix $\mathbf{P}_{(r)}$ and vector $(\overline{\boldsymbol{\alpha}}_n, \mathbf{0})$ given in Equation (4.29). The proof of it is made by induction, so in the next lemma is shown the first step of induction.

**Lemma 4.7** *The following properties hold using the MG-representations* $(\boldsymbol{\alpha}_j, \mathbf{S}_j, \mathbf{s}_j)$, *where* $\mathbf{s}_j = \mathbf{e} - \mathbf{S}_j \mathbf{e}$, $j = 1, \ldots, n$, *in Equations* (4.20) *and* (4.25).

*(a)* $Y_{(1:n)} \sim \mathrm{MG}\left(\overline{\boldsymbol{\alpha}}_n, \mathbf{C}_{(0,0)}, \mathbf{c}_{(0)}\right)$, *where* $\overline{\boldsymbol{\alpha}}_n = \boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n$ *and* $\mathbf{c}_{(0)} = \mathbf{e} - \mathbf{C}_{(0,0)}\mathbf{e}$.

*(b)* $\mathbf{C}_{(0,1,m)} = \mathbf{B}_{(0,1,m)}$.

*(c)* $\mathbf{C}_{(0,2,m)} = \mathbf{B}_{(0,2,m)}$.

**Proof.** For *(a)* consider Equation (4.24).

$$\begin{aligned}
\overline{\boldsymbol{\alpha}}_n \mathbf{C}_{(0,0,m)} \mathbf{e} &= (\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n)\left(\mathbf{S}_1^m \otimes \cdots \otimes \mathbf{S}_n^m\right)(\mathbf{e} \otimes \cdots \otimes \mathbf{e}) \\
&= (\boldsymbol{\alpha}_1 \mathbf{S}_1^m \mathbf{e}) \cdots (\boldsymbol{\alpha}_n \mathbf{S}_n^m \mathbf{e}) \quad \text{( by the Kronecker product property (A.6))} \\
&= \mathbb{P}\left(Y_1 > m, \ldots, Y_n > m\right) \quad \text{( by independence )}
\end{aligned}$$

$$= \quad \mathbb{P}\left(Y_{(1:n)} > m\right).$$

For *(b)* we can calculate the block matrix $\mathbf{C}_{(0,1,m)}$ as follows

$$\mathbf{C}_{(0,1,m)} = \sum_{k=0}^{m-1} \mathbf{C}_{(0,0,k)} \mathbf{C}_{(0,1)} \mathbf{C}_{(1,1,m-1-k)}, \tag{4.41}$$

(see Equation (4.25)). One entry of $\mathbf{C}_{(0,1,m)}$ is obtained by taking one specific 1-combination (this is one combination of one element of the set $\{1, 2, \ldots, n\}$). Let us take $(1)$. Then, the entry in $\mathbf{C}_{(0,1)}$ formed by the chosen 1-combination is

$$\mathbf{s}_1 \otimes \mathbf{S}_2 \otimes \cdots \otimes \mathbf{S}_n$$

(see Equation (4.21), and the entry in $\mathbf{C}_{(1,1,m-1-k)}$ is

$$\mathbf{S}_2^{m-1-k} \otimes \cdots \otimes \mathbf{S}_n^{m-1-k}.$$

Notice that (independently of the chosen 1-combination) we have

$$\mathbf{C}_{(0,0,k)} = \mathbf{S}_1^k \otimes \mathbf{S}_2^k \otimes \cdots \otimes \mathbf{S}_n^k, \tag{4.42}$$

(see Equation (4.24)).

Then by making product $\mathbf{C}_{(0,0,k)}\mathbf{C}_{(0,1)}$ we get

$$\mathbf{S}_1^k \mathbf{s}_1 \otimes \mathbf{S}_2^{k+1} \otimes \cdots \otimes \mathbf{S}_n^{k+1},$$

and now by multiplying with $\mathbf{C}_{(1,1,m-1-k)}$ we have

$$\mathbf{S}_1^k \mathbf{s}_1 \otimes \mathbf{S}_2^m \otimes \cdots \otimes \mathbf{S}_n^m.$$

The last operation is to take the sum in Equation (4.41), so we obtain

$$\sum_{k=0}^{m-1} \mathbf{S}_1^k \mathbf{s}_1 \otimes \mathbf{S}_2^m \otimes \cdots \otimes \mathbf{S}_n^m \quad = \sum_{(k) \in \mathcal{P}^{rep}_{1,(0,m-1)}} \mathbf{S}_1^k \mathbf{s}_1 \otimes \mathbf{S}_2^m \otimes \cdots \otimes \mathbf{S}_n^m. \tag{4.43}$$

Observe that the term in (4.43) is exactly the first entry in $\mathbf{B}_{(0,1,m)}$ (see Equation(4.40)), also, we get the same calculations for every 1-combination $(i)$, $i = 1, 2, \ldots, n$, and by using the lexicographical ordering we conclude that $\mathbf{C}_{(0,1,m)} = \mathbf{B}_{(0,1,m)}$.

For *(c)* we recall that $\mathbf{C}_{(0,2,m)}$ can be calculated as

$$\sum_{k=0}^{m-1} \left\{ \mathbf{C}_{(0,0,k)} \mathbf{C}_{(0,2)} \mathbf{C}_{(2,2,m-1-k)} + \sum_{j=0}^{k-1} \mathbf{C}_{(0,0,j)} \mathbf{C}_{(0,1)} \mathbf{C}_{(1,1,k-1-j)} \mathbf{C}_{(1,2)} \mathbf{C}_{(2,2,m-1-k)} \right\}, \tag{4.44}$$

(see Equation (4.25)).

We are going to obtain the first entry of $\mathbf{C}_{(0,2,m)}$, and for that we choose to take the 2-combination given by $\{1,2\}$ (due to the lexicographical ordering).

In Equation (4.44) we can see two products:

$$\mathbf{C}_{(0,0,k)}\mathbf{C}_{(0,2)}\mathbf{C}_{(2,2,m-1-k)} \tag{4.45}$$

and

$$\mathbf{C}_{(0,0,j)}\mathbf{C}_{(0,1)}\mathbf{C}_{(1,1,k-1-j)}\mathbf{C}_{(1,2)}\mathbf{C}_{(2,2m-1-k)}. \tag{4.46}$$

In order to calculate the expression in Equation (4.45) we use the 2-combination given by $\{1,2\}$. The block $\mathbf{C}_{(0,0,k)}$ is given in Equation (4.42) (and it is independent of the chosen 2-combination). The corresponding block entry in $\mathbf{C}_{(0,2)}$ is

$$\mathbf{s}_1 \otimes \mathbf{s}_2 \otimes \mathbf{S}_3 \otimes \cdots \otimes \mathbf{S}_n$$

(see Equation (4.21)), and the corresponding block entry in $\mathbf{C}_{(2,2,m-1-k)}$ is

$$\mathbf{S}_3^{m-1-k} \otimes \cdots \otimes \mathbf{S}_n^{m-1-k}$$

(see Equation (4.24)).

For the product $\mathbf{C}_{(0,0,k)}\mathbf{C}_{(0,2)}$ we multiply the entries $\mathbf{S}_1^k \otimes \mathbf{S}_2^k \otimes \cdots \otimes \mathbf{S}_n^k$ from $\mathbf{C}_{(0,0,k)}$ and $\mathbf{s}_1 \otimes \mathbf{s}_2 \otimes \mathbf{S}_3 \otimes \cdots \otimes \mathbf{S}_n$ from $\mathbf{C}_{(0,2)}$, and it gives the term

$$\mathbf{S}_1^k\mathbf{s}_1 \otimes \mathbf{S}_2^k\mathbf{s}_2 \otimes \mathbf{S}_3^{k+1} \otimes \cdots \otimes \mathbf{S}_n^{k+1}. \tag{4.47}$$

Then, by multiplying the term $\mathbf{S}_3^{m-1-k} \otimes \cdots \otimes \mathbf{S}_n^{m-1-k}$ from $\mathbf{C}_{(2,2,m-1-k)}$ with the term in (4.47), we get

$$\mathbf{S}_1^k\mathbf{s}_1 \otimes \mathbf{S}_2^k\mathbf{s}_2 \otimes \mathbf{S}_3^{m} \otimes \cdots \otimes \mathbf{S}_n^{m}. \tag{4.48}$$

Thus, the term (4.48) is given by choosing the 2-combination $\{1,2\}$ and making the product $\mathbf{C}_{(0,0,k)}\mathbf{C}_{(0,2)}\mathbf{C}_{(2,2,m-1-k)}$.

Now, to calculate the expression in (4.46) we are going to use the chosen 2-combination $\{1,2\}$ again. First, we have

$$\mathbf{C}_{(0,0,j)} = \mathbf{S}_1^j \otimes \mathbf{S}_2^j \otimes \mathbf{S}_3^j \otimes \cdots \otimes \mathbf{S}_n^j, \tag{4.49}$$

(see Equation (4.42)).

For the block matrix $\mathbf{C}_{(0,1)}$, we split the chosen combination so that we obtain 1-combinations. In this case we have two 1-combinations: $\{1\}$ and $\{2\}$. The corresponding entries in $\mathbf{C}_{(0,1)}$ for the 1-combinations are given next.

$$\mathbf{s}_1 \otimes \mathbf{S}_2 \otimes \mathbf{S}_3 \otimes \cdots \otimes \mathbf{S}_n \quad \text{for the 1-combination } \{1\}, \tag{4.50}$$

$$\mathbf{S}_1 \otimes \mathbf{s}_2 \otimes \mathbf{S}_3 \otimes \cdots \otimes \mathbf{S}_n \quad \text{for the 1-combination } \{2\}, \tag{4.51}$$

(see Equation (4.21)).

We also take the split of the chosen combination to obtain the entries in the block matrix $\mathbf{C}_{(1,1,k-1-j)}$, so we get

$$\mathbf{S}_2^{k-1-j} \otimes \mathbf{S}_3^{k-1-j} \otimes \cdots \otimes \mathbf{S}_n^{k-1-j} \quad \text{for the 1-combination } \{1\}, \tag{4.52}$$

$$\mathbf{S}_1^{k-1-j} \otimes \mathbf{S}_3^{k-1-j} \otimes \cdots \otimes \mathbf{S}_n^{k-1-j} \quad \text{for the 1-combination } \{2\}, \tag{4.53}$$

(see Equation (4.24)).

Now we have two cases to calculate the corresponding block entries in $\mathbf{C}_{(1,2)}$.

**Case 1** : The 1-combination $\{1\}$ and the 2-combination $\{1,2\}$ form the Kronecker product:

$$\mathbf{s}_2 \otimes \mathbf{S}_3 \otimes \cdots \otimes \mathbf{S}_n, \tag{4.54}$$

see Equation (4.23).

**Case 2** : The 1-combination $\{2\}$ and the 2-combination $\{1,2\}$ form the Kronecker product:

$$\mathbf{s}_1 \otimes \mathbf{S}_3 \otimes \cdots \otimes \mathbf{S}_n, \tag{4.55}$$

see Equation (4.23).

Finally, the corresponding block entry in $\mathbf{C}_{(2,2,m-1-k)}$ that comes from the 2-combination $\{1,2\}$ is

$$\mathbf{S}_3^{m-1-k} \otimes \cdots \otimes \mathbf{S}_n^{m-1-k}, \tag{4.56}$$

see Equation (4.24).

Until here we have recognized the corresponding entries and we have observed that there are two cases, so the expression in (4.46) is going to be calculated in the two cases.

**For case 1**: The product of (4.49) and (4.50) is

$$\mathbf{S}_1^{j}\mathbf{s}_1 \otimes \mathbf{S}_2^{j+1} \otimes \mathbf{S}_3^{j+1} \otimes \cdots \otimes \mathbf{S}_n^{j+1}. \tag{4.57}$$

Then, the product of (4.57) with (4.52) is

$$\mathbf{S}_1^j\mathbf{s}_1 \otimes \mathbf{S}_2^k \otimes \mathbf{S}_3^k \otimes \cdots \otimes \mathbf{S}_n^k. \tag{4.58}$$

The product of (4.58) with (4.54) is

$$\mathbf{S}_1^j\mathbf{s}_1 \otimes \mathbf{S}_2^k\mathbf{s}_2 \otimes \mathbf{S}_3^{k+1} \otimes \cdots \otimes \mathbf{S}_n^{k+1}. \tag{4.59}$$

By calculating the product of (4.59) with (4.56), we obtain

$$\mathbf{S}_1^j\mathbf{s}_1 \otimes \mathbf{S}_2^k\mathbf{s}_2 \otimes \mathbf{S}_3^m \otimes \cdots \otimes \mathbf{S}_n^m. \tag{4.60}$$

**For case 2**: We basically redo the steps made for case 1 but now by using the entries for case 2.

The product of (4.49) and (4.51) is

$$\mathbf{S}_1^{j+1} \otimes \mathbf{S}_2^j\mathbf{s}_2 \otimes \mathbf{S}_3^{j+1} \otimes \cdots \otimes \mathbf{S}_n^{j+1}. \tag{4.61}$$

The product of expression (4.61) with (4.53) is

$$\mathbf{S}_1^k \otimes \mathbf{S}_2^j\mathbf{s}_2 \otimes \mathbf{S}_3^k \otimes \cdots \otimes \mathbf{S}_n^k. \tag{4.62}$$

The product of expression (4.62) with (4.55) is

$$\mathbf{S}_1^k\mathbf{s}_1 \otimes \mathbf{S}_2^j\mathbf{s}_2 \otimes \mathbf{S}_3^{k+1} \otimes \cdots \otimes \mathbf{S}_n^{k+1}. \tag{4.63}$$

By calculating the product of expression (4.63) with (4.56), we obtain

$$\mathbf{S}_1^k\mathbf{s}_1 \otimes \mathbf{S}_2^j\mathbf{s}_2 \otimes \mathbf{S}_3^m \otimes \cdots \otimes \mathbf{S}_n^m. \tag{4.64}$$

By collecting the resulting terms in (4.48), (4.60) and (4.64) and taking the sum in (4.44), we finally obtain the first block entry of $\mathbf{C}_{(0,2,m)}$. This is given by

$$\begin{aligned}&\sum_{k=0}^{m-1}\Bigg\{\mathbf{S}_1^k\mathbf{s}_1 \otimes \mathbf{S}_2^k\mathbf{s}_2 \otimes \mathbf{S}_3^m \otimes \cdots \otimes \mathbf{S}_n^m + \\ &\sum_{j=0}^{k-1}\left\{\mathbf{S}_1^j\mathbf{s}_1 \otimes \mathbf{S}_2^k\mathbf{s}_2 \otimes \mathbf{S}_3^m \otimes \cdots \otimes \mathbf{S}_n^m + \mathbf{S}_1^k\mathbf{s}_1 \otimes \mathbf{S}_2^j\mathbf{s}_2 \otimes \mathbf{S}_3^m \otimes \cdots \otimes \mathbf{S}_n^m\right\}\Bigg\}.\end{aligned} \tag{4.65}$$

The number of the resulting couple of powers $(k_1, k_2)$ from the last sum corresponds to the number of permutations (with repetitions) of size two from the set $\{0, 1, \ldots, m-1\}$ and it coincides with the number of terms of the type

$$\mathbf{S}_1^{k_1}\mathbf{s}_1 \otimes \mathbf{S}_2^{k_2}\mathbf{s}_2 \otimes \mathbf{S}_3^m \otimes \cdots \otimes \mathbf{S}_n^m.$$

Thus, we conclude that $\mathbf{C}_{(0,2,m)} = \mathbf{B}_{(0,2,m)}$ (see Equation (4.40)). $\square$

**Lemma 4.8** *For every $i = 1, \ldots, n-1$, and $m \in \mathbb{N}$, we have $\mathbf{C}_{(0,i,m)} = \mathbf{B}_{(0,i,m)}$, where $\mathbf{C}_{(0,i,m)}$ is the block matrix defined as in Equation (4.25) formed by the MG-representations $(\boldsymbol{\alpha}_j, \mathbf{S}_j, \mathbf{s}_j)$, where $\mathbf{s}_j = \mathbf{e} - \mathbf{S}_j\mathbf{e}$, for all $j = 1, \ldots, n$.*

**Proof.** We are going to prove this lemma by making induction only over the first block entry of $\mathbf{C}_{(0,i,m)}$, since for any other block entry we have the same calculations.

For $i = 1, 2$, we already calculate the first block entry in $\mathbf{C}_{(0,i,m)}$, this is given in expression (4.43) (see Lemma (4.7)).

The induction hypothesis is going to be applied for a fixed $i$, where $i = 1, \ldots, n-1$, and it is also valid for all $j < i$. We choose the $i$-combination given by $\big(1, 2, \ldots, i\big)$, which corresponds to the first block entry of $\mathbf{C}_{(0,i,m)}$ due to the lexicographical ordering. Then, we assume that the first block entry of $\mathbf{C}_{(0,i,m)}$ is given by

$$\sum_{(k_1,k_2,\ldots,k_i)\in\mathcal{P}^{rep}_{i,(0,m-1)}} \mathbf{S}_1^{k_1}\mathbf{s}_1 \otimes \mathbf{S}_2^{k_2}\mathbf{s}_2 \otimes \cdots \otimes \mathbf{S}_i^{k_i}\mathbf{s}_i \otimes \mathbf{S}_{i+1}^m \otimes \cdots \otimes \mathbf{S}_n^m, \tag{4.66}$$

this is the induction hypothesis.

We recall Equation (4.25), this is given by

$$\mathbf{C}_{(0,i+1,m)} = \sum_{k=0}^{m-1} \left\{ \mathbf{C}_{(0,0,k)}\mathbf{C}_{(0,i+1)} + \sum_{l=1}^{i} \mathbf{C}_{(0,l,k)}\mathbf{C}_{(l,i+1)} \right\} \mathbf{C}_{(i+1,i+1,m-1-k)}, \tag{4.67}$$

and consider the $i+1$-combination $\big(1, 2, \ldots, i, i+1\big)$.

Now, observe that from the next expression

$$\mathbf{C}_{(0,l,k)}\mathbf{C}_{(l,i+1)}\mathbf{C}_{(i+1,i+1,m-1-k)},$$

where $k \in \big\{0, 1, \ldots, m-1\big\}$ is fixed, we get terms of the type

$$\mathbf{S}_1^{k_1}\mathbf{s}_1 \otimes \mathbf{S}_2^{k_2}\mathbf{s}_2 \otimes \cdots \mathbf{S}_l^{k_l}\mathbf{s}_l \otimes \mathbf{S}_{l+1}^{k}\mathbf{s}_{l+1} \otimes \cdots \mathbf{S}_{i+1}^{k}\mathbf{s}_{i+1} \otimes \mathbf{S}_{i+2}^m \otimes \cdots \otimes \mathbf{S}_n^m. \tag{4.68}$$

Here, the powers $(k_1, k_2, \ldots, k_l)$ form a set of permutations (with repetitions) of the size $l$ from the set $\{0, 1, \ldots, k-1\}$ (this is by the induction hypothesis in Equation (4.66). Now, by adding the powers $(k, k, \ldots, k)$ of size $i+1-l$ to every permutation of powers of size $l$ in the following way

$$(k_1, k_2, \ldots, k_l, k, k \ldots, k),$$

we then get values of powers of size $i+1$ that correspond to values of powers obtained from the expression in (4.68).

Now, notice that in the expression $\mathbf{C}_{(0,0,k)}\mathbf{C}_{(0,i+1)}\mathbf{C}_{(i+1,i+1,m-1-k)}$ we have terms of the type

$$\mathbf{S}_1^{k}\mathbf{s}_1 \otimes \mathbf{S}_2^{k}\mathbf{s}_2 \otimes \cdots \otimes \mathbf{S}_{i+1}^{k}\mathbf{s}_{i+1} \otimes \mathbf{S}_{i+2}^m \otimes \cdots \otimes \mathbf{S}_n^m,$$

and the powers $(k, k, \ldots, k)$ form a set of repetitions of $\{k\}$ of size $i + 1$.

For a fixed $k \in \{0, 1, \ldots, m - 1\}$ we can count all the Kronecker products obtained from every product of block matrices in Equation (4.67). The number of all Kronecker products is given by

$$\sum_{l=0}^{i} \binom{i+1}{l} \cdot k^l.$$

Finally, since the value of $k$ is taken from 0 to $m - 1$, then

$$\begin{aligned} \sum_{k=0}^{m-1} \sum_{l=0}^{i} \binom{i+1}{l} \cdot k^l &= 1 + \sum_{l=0}^{i} \binom{i+1}{l} \left( \sum_{k=1}^{m-1} k^l \right) \\ &= 1 + (m^{i+1} - 1) \quad \text{(by Pascal's identity)} \\ &= m^{i+1}, \end{aligned}$$

which means that in Equation (4.67) are in total $m^{i+1}$ Kronecker products of the type

$$\mathbf{S}_1^{k_1} \mathbf{s}_1 \otimes \mathbf{S}_2^{k_2} \mathbf{s}_2 \otimes \cdots \otimes \mathbf{S}_{i+1}^{k_{i+1}} \mathbf{s}_{i+1} \otimes \mathbf{S}_{i+2}^{m} \otimes \cdots \otimes \mathbf{S}_n^{m}.$$

Now we conclude that the first block entry of $\mathbf{C}_{(0,i+1,m)}$ is given by

$$\sum_{(k_1,k_2,\ldots,k_{i+1}) \in \mathcal{P}^{rep}_{i+1,(0,m-1)}} \mathbf{S}_1^{k_1} \mathbf{s}_1 \otimes \mathbf{S}_2^{k_2} \mathbf{s}_2 \otimes \cdots \otimes \mathbf{S}_{i+1}^{k_{i+1}} \mathbf{s}_{i+1} \otimes \mathbf{S}_{i+2}^{m} \otimes \cdots \otimes \mathbf{S}_n^{m},$$

which is exactly the first block entry of $\mathbf{B}_{(0,i+1,m)}$. □

**Theorem 4.9** *The distribution of $Y_{(r:n)}$ has a MG-representation given by*

$$\left( (\overline{\boldsymbol{\alpha}}_n, \mathbf{0}), \mathbf{P}_{(r)}, \mathbf{p}_{(r)} \right),$$

*where $\overline{\boldsymbol{\alpha}}_n = \boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n$, $\mathbf{P}_{(r)}$ is defined as in the Equation* (4.28) *and $\mathbf{p}_{(r)} = \mathbf{e} - \mathbf{P}_{(r)}\mathbf{e}$.*

**Proof.** It follows directly from Lemma 4.8. □

### 4.5.2 Second type of MG-representations

Here, we present other type of representation for the order statistics which are derived directly from the formula of the survival function of order statistics given in Equation (4.2).

Let $Y_1, Y_2, \ldots, Y_n$ be a set of independent with distribution given by $Y_j \sim \mathrm{MG}(\boldsymbol{\alpha}_j, \mathbf{S}_j, \mathbf{s}_j)$, $j = 1, \ldots, n$, where $\mathbf{s}_j = \mathbf{e} - \mathbf{S}_j\mathbf{e}$, for all $j = 1, \ldots, n$. Let $F_j(m)$ denote the cumulative distribution function of $Y_j$, $j = 1, \ldots, n$, respectively.

Let $\overline{F}_i(m) = 1 - F_i(m)$ for all $i = 1, \ldots, n$. Thus, Equation (4.2) can be written in terms of $\overline{F}_i(m)$ as follows

$$\mathbb{P}\left(Y_{(r:n)} > m\right) = \sum_{i=0}^{r-1} \frac{1}{i!(n-i)!} \mathrm{per}\big[\underbrace{\mathbf{e} - \overline{\mathbf{F}}}_{\substack{i \\ \text{columns}}} \quad \underbrace{\overline{\mathbf{F}}}_{\substack{n-i \\ \text{columns}}}\big], \tag{4.69}$$

where $\overline{\mathbf{F}} = \left(\overline{F}_1(m), \ldots, \overline{F}_n(m)\right)^\top$.

Observe that one product from the $i$ columns of $\mathbf{e} - \overline{\mathbf{F}}$ is

$$\left(1 - \overline{F}_1(m)\right)\left(1 - \overline{F}_2(m)\right) \cdots \left(1 - \overline{F}_i(m)\right) \quad = \quad \sum_{k=0}^{i} (-1)^k \Big( \sum_{\boldsymbol{\gamma} \in \mathcal{C}_{k,i}} \prod_{\ell \in \boldsymbol{\gamma}} \overline{F}_\ell(m)\Big),$$

where $\sum\limits_{\boldsymbol{\gamma} \in \mathcal{C}_{0,i}} \prod\limits_{\ell \in \boldsymbol{\gamma}} \overline{F}_\ell(m) = 1.$

Now let us introduce the next matrices

$$\mathbf{A}_{n-i} = \begin{bmatrix} \underbrace{\mathbf{e}}_{\substack{i \\ \text{columns}}} & \underbrace{\overline{\mathbf{F}}}_{\substack{n-i \\ \text{columns}}} \end{bmatrix}, \quad i = 0, 1 \ldots, n.$$

Since the permanent based formula is invariant under arbitrary permutations of columns, then we have the following relation

$$\mathrm{per}\begin{bmatrix} \underbrace{\mathbf{e} - \overline{\mathbf{F}}}_{\substack{i \\ \text{columns}}} & \underbrace{\overline{\mathbf{F}}}_{\substack{n-i \\ \text{columns}}} \end{bmatrix} = \sum_{k=0}^{i} (-1)^k \binom{i}{k} \mathrm{per}\left[\mathbf{A}_{n-i+k}\right]. \tag{4.70}$$

By combining Equations (4.69) and (4.70), we get the following calculations

$$\begin{aligned}
\mathbb{P}\left(Y_{(r:n)} > m\right) &= \sum_{i=0}^{r-1} \sum_{k=0}^{i} \frac{(-1)^k \binom{i}{k}}{i!(n-i)!} \mathrm{per}\left[\mathbf{A}_{n-i+k}\right] \\
&= \sum_{i=0}^{r-1} \sum_{l=0}^{i} \frac{(-1)^{i-l}}{l!(i-l)!(n-i)!} \mathrm{per}\left[\mathbf{A}_{n-l}\right], \quad l = i-k, \\
&= \sum_{\ell=0}^{r-1} \left( \sum_{i=\ell}^{r-1} \frac{(-1)^{i-\ell}}{\ell!(i-\ell)!(n-i)!} \right) \mathrm{per}\left[\mathbf{A}_{n-\ell}\right].
\end{aligned} \tag{4.71}$$

For every matrix $\mathbf{C}_{(i,i)}$, $i = 0, \ldots, n-1$, (see Equation (4.22)), we define the vector $\boldsymbol{\beta}_i$ as follows

$$\boldsymbol{\beta}_i = \left\{ \bigotimes_{\jmath=1}^{n} g_\jmath^\gamma : \quad \gamma \in \mathcal{C}_{i,n} \right\}, \quad g_\jmath^\gamma = \begin{cases} 1 & \text{if } \jmath \in \gamma \\ \boldsymbol{\alpha}_\jmath & \text{if } \jmath \notin \gamma \end{cases},$$

where the entries are placed consecutively and following the lexicographical ordering. Hence we have

$$\boldsymbol{\beta}_i \mathbf{C}_{(i,i)}^m \mathbf{e} = \sum_{\gamma \in \mathcal{C}_{i,n}} \left( \bigotimes_{\imath \notin \gamma} \boldsymbol{\alpha}_\imath \right) \left( \bigotimes_{\imath \notin \gamma} \mathbf{S}_\imath^m \right) \left( \bigotimes_{\imath \notin \gamma} \mathbf{e}_\imath \right),$$

and consequently

$$\text{per}\left[\mathbf{A}_{n-i}\right] = i!(n-i)!\boldsymbol{\beta}_i \mathbf{C}_{(i,i)}^m \mathbf{e}, \tag{4.72}$$

for every $i = 0, 1, \ldots, n-1$.

By replacing Equation (4.72) into Equation (4.71) we get

$$\mathbb{P}\left(Y_{(r:n)} > m\right) = \sum_{\ell=0}^{r-1} \left( \sum_{i=\ell}^{r-1} \frac{(-1)^{i-\ell}}{(i-\ell)!(n-i)!} \right) (n-\ell)!\boldsymbol{\beta}_\ell \mathbf{C}_{(\ell,\ell)}^m \mathbf{e}. \tag{4.73}$$

Denote

$$\boldsymbol{\pi}_\ell = \left( \sum_{i=\ell}^{r-1} \frac{(-1)^{i-\ell}(n-\ell)!}{(i-\ell)!(n-i)!} \right) \boldsymbol{\beta}_\ell, \quad \ell = 0, \ldots, r-1. \tag{4.74}$$

Thus we can write Equation (4.73) as follows

$$\begin{aligned} \mathbb{P}\left(Y_{(r:n)} > m\right) &= \sum_{\ell=0}^{r-1} \boldsymbol{\pi}_\ell \mathbf{C}_{(\ell,\ell)}^m \mathbf{e} \\ &= \left(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{r-1}\right) \begin{pmatrix} \mathbf{C}_{(0,0)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{(1,1)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots\vdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{(r-1,r-1)} \end{pmatrix}^m \mathbf{e}. \end{aligned}$$

Finally we are able to conclude with the following Theorem.

**Theorem 4.10** *Let $Y_1, Y_2, \ldots, Y_n$ be independent and MG-distributed random variables (ME-distributed) with representations given by $\left(\boldsymbol{\alpha}_j, \mathbf{S}_j, \mathbf{s}_j\right)$, where $\mathbf{s}_j = \mathbf{e} - \mathbf{S}_j \mathbf{e}$, (or $\mathbf{s}_j = -\mathbf{S}_j \mathbf{e}$), $j = 1, \ldots, n$, respectively. Then, $Y_{(r:n)}$ has a MG-representation (or ME-representation) given by $\left(\boldsymbol{\pi}_{(r)}, \mathbf{C}_{(r)}, \mathbf{c}_{(r)}\right)$, where*

$$\boldsymbol{\pi}_{(r)} = \left(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_{r-1}\right),$$

$\boldsymbol{\pi}_\ell$ *is defined in Equation* (4.74),

$$\mathbf{C}_{(r)} = diag\Big\{\mathbf{C}_{(\ell,\ell)} : \quad \ell = 0, \ldots, r-1.\Big\},$$

$\mathbf{C}_{(\ell,\ell)}$ *is defined as in Equation* (4.22)*, and*

$$\mathbf{c}_{(r)} = \mathbf{e} - \mathbf{C}_{(r)}\mathbf{e},$$

$\big($*or* $\mathbf{c}_{(r)} = -\mathbf{C}_{(r)}\mathbf{e}$ *in the case of ME-distributions*$\big)$.

**Corollary 4.11** *Let* $Y_1, Y_2, \ldots, Y_n$ *be independent and identically MG-distributed random variables with representation* $\mathrm{MG}(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s})$, *where* $\mathbf{s} = \mathbf{e} - \mathbf{S}\mathbf{e}$. *Let* $\left((\overline{\boldsymbol{\alpha}}_n, \mathbf{0}), \mathbf{P}_{(r)}\right)$ *be the same representation considered in Section 4.4. Then, the couple* $\left(\mathbf{q}_{(r)}, \mathbf{Q}_{(r)}\right)$, *(defined in the Section* 4.4.1*) is another representation for the* $r$*-th order statistic* $Y_{(r:n)}$.

The proof is exactly the same than the proof presented in Corollary 4.6.

## 4.6 Concluding remarks

We have seen that we can apply the same representations for order statistics and which come from a probabilistic interpretation to the case of matrix-geometric or matrix-exponential distributions and also that we have considered an order reduction for those representations. Furthermore, we have shown another type of representation for order statistics without a probabilistic interpretation and which can be applied to the representations for either order statistics from matrix-geometric distributions or order statistics from matrix-exponential distributions.

The proof for every type of representation is given by making direct calculations with an exhaustive use of the Kronecker products. Therefore, we are open to explore other ways to provide a more practical and elegant proof for the given representations.

Other aspects of the given representations that may be convenient to analyse are for example: the non-singularity of representations, other methods or techniques for order reduction of the representations and the problem of the minimal order of these representations. For instance, we have that the second type of representations given in Theorem 4.10 consists on a diagonal block matrix of Kronecker products and due to the Kronecker product is nonsingular if and only if the matrices are nonsingular, then we conclude that the representation is nonsingular under the assumption of all the representations of the random variables are nonsingular.

In [Abd11] is given a method which is based on Laplace transforms to provide the moments of order statistics from phase-type distributions. This method does not have any restriction to be applied for the moments of order statistics from Matrix-Exponential distributions.

For the order reduction of representations, in [HA17] is shown an algorithm for constructing representations for order statistics of independent and identically distributed discrete phase-type and phase-type distributions, where the new dimension of the representations can be reduced considerably.

CHAPTER 5

# Multivariate Discrete Phase-type distributions (MDPH*)

In this chapter we present our study of the discrete version of the class of multivariate distributions of phase-type provided by V. G. Kulkarni in [Kul89]. The class of multivariate discrete distributions that we consider here is denoted by MDPH*.

Our aim is to create an unified approach (parallel to the analysis of the class of distributions in continuous time) to develop theoretical results in the discrete case such as the joint probability-generating function, joint moment-generating function, closure properties and several examples of representations of this distribution.

The chapter starts with a section that explains the construction of the multivariate discrete phase-type which is an equivalent to the construction of the continuous case. Then, we present the section of the class of MDPH*-distributions where we introduce the formal definition and many properties of the distribution such as the distribution of the marginals, a closed form formula for the joint probability-generating function and moments, as well as some closure properties of the class. In the section of Examples, we try to cover many examples which are found in the literature and we show their corresponding MDPH*-representation. Some examples are well-known since long ago, such as the Bivariate geometric distribution and Bivariate Negative Binomial distribution, and we also present some more recent multivariate discrete phase-type distributions such as the compound multivariate phase-type distributions.

## 5.1 Construction of the MDPH* class

### 5.1.1 The univariate case.

Let us consider $\tau \sim \text{DPH}_p(\boldsymbol{\pi}, \mathbf{T})$ and let $\{X_m\}_{m \in \mathbb{N}}$ be its underlying Markov chain with state space

$$\mathcal{E} = \{1, 2, \ldots, p, p+1\},$$

where $p+1$ is the absorbing state and the rest of the states are transient.

Let

$$Y = \sum_{m=0}^{\tau-1} r(X_m)$$

where $r(\cdot)$ is a function from the state-space $\mathcal{E} - \{p+1\}$ to $\mathbb{N}_0^p$. The values of the function $r(\cdot)$ are called "rewards".

For the construction of the MDPH* class, first we prove that $Y$ is DPH-distributed and then we proceed with the multivariate case.

In the next section, we work with the case where the rewards are 0 or 1.

#### 5.1.1.1 0-1 valued rewards

Let $r(\cdot)$ be a function from $\mathcal{E} - \{p+1\}$ to $\{0, 1\}$ and consider the vector

$$\mathbf{r} = (r(1), ..., r(p)),$$

which is going to be referred as the "vector of rewards".

Since $r(i)$ is either zero or one (for every $i = 1, \ldots, p$), then the subset $\mathcal{E} - \{p+1\}$ can be partitioned into the subsets $\mathcal{E}_+$ and $\mathcal{E}_0$, where

$$\mathcal{E}_+ = \Big\{ i \in \mathcal{E} : r(i) = 1 \Big\} \quad \text{and} \quad \mathcal{E}_0 = \Big\{ i \in \mathcal{E} : r(i) = 0 \Big\}.$$

The initial distribution $\boldsymbol{\pi}$ and the sub-transition probability matrix $\mathbf{T}$ can be partitioned according to the subsets $\mathcal{E}_+$ and $\mathcal{E}_0$ as follows

$$\boldsymbol{\pi} = (\boldsymbol{\pi}_+, \boldsymbol{\pi}_0) \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} \mathbf{T}_{++} & \mathbf{T}_{+0} \\ \mathbf{T}_{0+} & \mathbf{T}_{00} \end{pmatrix}.$$

Recall the random variable

$$Y = \sum_{m=0}^{\tau-1} r(X_m)$$

and notice that $Y$ represents the total time the Markov chain $\{X_m\}_{m\in\mathbb{N}}$ spends in the set $\mathcal{E}_+$ prior to absorption. According to this interpretation we state the following Lemma.

**Lemma 5.1** *The distribution of*

$$Y = \sum_{m=0}^{\tau-1} r(X_m)$$

*is a mixture of a zero point and a discrete phase–type distribution with representation* $(\boldsymbol{\pi}^*, \mathbf{T}^*)$*, where*

$$\begin{aligned} \boldsymbol{\pi}^* &= \boldsymbol{\pi}_+ + \boldsymbol{\pi}_0(\mathbf{I} - \mathbf{T}_{00})^{-1}\mathbf{T}_{0+}, \\ \mathbf{T}^* &= \mathbf{T}_{++} + \mathbf{T}_{+0}(\mathbf{I} - \mathbf{T}_{00})^{-1}\mathbf{T}_{0+}. \end{aligned}$$

*The zero point is of size* $1 - \boldsymbol{\pi}^*\mathbf{e}$.

**Proof.**

The idea of the proof is to construct an absorbing Markov chain from the original one such that the distribution of its time to absorption coincides with the distribution of $Y$.

Let $\{X^*_m\}_{m\in\mathbb{N}}$ denote the new Markov chain with state space given by $\mathcal{E}_+ \cup \{p+1\}$. This new Markov chain will only consider the transitions of $\{X_m\}_{m\in\mathbb{N}}$ in the subset of states $\mathcal{E}_+$ before it gets absorbed.

First we are going to derive the initial distribution of $\{X^*_m\}_{m\in\mathbb{N}}$. For that purpose we consider two possibilities:

(1) $\{X_m\}_{m\in\mathbb{N}}$ starts in a state in $\mathcal{E}_+$.

(2) $\{X_m\}_{m\in\mathbb{N}}$ starts in a state in $\mathcal{E}_0$ and in a some finite time it jumps to $\mathcal{E}_+$.

In the case (1), the probability of $\{X_m\}_{m\in\mathbb{N}}$ starting in state $i \in \mathcal{E}_+$ is denoted by $\pi_{+,i}$. In the case (2), the probability of $\{X_m\}_{m\in\mathbb{N}}$ initiating in state $j \in \mathcal{E}_0$ and then it jumps to $i \in \mathcal{E}_+$ at some (finite) time is given by

$$\mathbf{e}_j^\top \sum_{n=0}^{\infty} (\mathbf{T}_{00})^n \, \mathbf{T}_{0+}\mathbf{e}_i = \mathbf{e}_j^\top \left(\mathbf{I} - \mathbf{T}_{00}\right)^{-1} \mathbf{T}_{0+}\mathbf{e}_i,$$

where the vector $\mathbf{e}_j^\top$ denotes a row vector which its $j$-th entry is equal to 1 and the rest of the entries are zero, while $\mathbf{e}_i$ denotes a column vector which its $i$-th entry is equal to 1 and the rest of its entries are equal to zero.

Then, for every $i \in \mathcal{E}_+$ we obtain

$$\mathbb{P}\left(X_0^* = i\right) = \pi_{+,i} + \boldsymbol{\pi}_0 \left(\mathbf{I} - \mathbf{T}_{00}\right)^{-1} \mathbf{T}_{0+}\mathbf{e}_i.$$

Let us denote $\boldsymbol{\pi}_i^* = \mathbb{P}\left(X_0^* = i\right)$ and let $\boldsymbol{\pi}^*$ denotes the row vector with entries given by $\boldsymbol{\pi}_i^*, i \in \mathcal{E}_+$. Then, the initial probability of $\{X_m^*\}_{m\in\mathbb{N}}$ is given by the vector

$$\left(\boldsymbol{\pi}^*, 1 - \boldsymbol{\pi}^*\mathbf{e}\right).$$

Now, in order to calculate the sub-transition probability matrix of $\{X_m^*\}_{m\in\mathbb{N}}$ , based on the new initial probability $\boldsymbol{\pi}^*$, we have the next two possibilities of transitions between states in $\mathcal{E} - \{p+1\}$ for $\{X_m\}_{m\in\mathbb{N}}$:

(1) $\{X_m\}_{m\in\mathbb{N}}$ goes from one state in $\mathcal{E}_+$ to another state in $\mathcal{E}_+$.

(2) $\{X_m\}_{m\in\mathbb{N}}$ goes from one state in $\mathcal{E}_+$ to another state in $\mathcal{E}_0$.

Therefore, the sub-transition probability matrix of $\{X_m^*\}_{m\in\mathbb{N}}$ is given by

$$\mathbf{T}^* = \mathbf{T}_{++} + \mathbf{T}_{+0}(\mathbf{I} - \mathbf{T}_{00})^{-1}\mathbf{T}_{0+}.$$

□

#### 5.1.1.2 Non-negative integer valued rewards

In this section, we allow the vector of rewards $\mathbf{r}$ to take values in $\mathbb{N}_0$ and we determine a DPH-representation for

$$Y = \sum_{m=0}^{\tau-1} r(X_m).$$

As in the case of 0-1-valued rewards, we are going to get a DPH-representation for $Y$ by constructing an appropriate absorbing Markov chain from the original one.

Let $\{X^\star_m\}_{m\in\mathbb{N}}$ denote the new Markov chain which is going to be constructed. Assume that at time $n \in \mathbb{N}$ the Markov chain $\{X_m\}_{m\in\mathbb{N}}$ is in the state $i \in \mathcal{E}_+$. Thus, $r(X_n) = r(i)$. Now, in order to get the value $r(i)$ from $\{X_m\}_{m\in\mathbb{N}}$, we add $r(i) - 1$ transient states more in the original state space $\mathcal{E}$ in the way that $\{X_m\}_{m\in\mathbb{N}}$ is forced to pass through all of them after it visits the state $i$. Thus, it takes $r(i) - 1$ more steps before $\{X_m\}_{m\in\mathbb{N}}$ gets absorbed. This idea is formulated next.

For every $i \in \mathcal{E}_+$, consider a new set of states given by

$$\mathbf{h}_i = \left\{i_1, i_2, \ldots, i_{r(i)}\right\},$$

where $i_1 = i$ and the rest of the states are the new transient states.

The state space for the new Markov chain $\{X^\star_m\}_{m\in\mathbb{N}}$ is given by

$$\mathcal{E}^\star = \mathcal{E}_0 \cup \left\{p+1\right\} \cup \left(\bigcup_{i\in\mathcal{E}_+} \mathbf{h}_i\right).$$

The initial distribution of $\{X^\star_m\}_{m\in\mathbb{N}}$ is denoted as

$$\widetilde{\boldsymbol{\pi}} = (\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_p),$$

where for every $i \in \mathcal{E}_+$

$$\boldsymbol{\pi}_i = \left(\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_{r(i)}}\right),$$

and $\pi_{i_1} = \mathbb{P}\left(X_0 = i\right)$ and $\pi_{i_l} = 0$ for all $l = 2, \ldots, r(i)$, while for every $i \in \mathcal{E}_0$, $\boldsymbol{\pi}_i = \mathbb{P}\left(X_0 = i\right)$.

The sub-transition probability matrix of $\{X^\star_m\}_{m\in\mathbb{N}}$ is going to be constructed by incorporating new blocks of matrices to the original sub-transition probability matrix.

Let $i, j \in \mathcal{E}_+$ and consider the sets of new states

$$\mathbf{h}_i = \left\{i_1, i_2, \ldots, i_{r(i)}\right\} \quad \text{and} \quad \mathbf{h}_j = \left\{j_1, j_2, \ldots, j_{r(j)}\right\}.$$

In the case where $i = j$, let $\widetilde{\mathbf{T}}_{ij}$ denote an $r(i)$-square matrix where every entry is defined as follows

$$\widetilde{t}_{\ell,k} = \begin{cases} 1 & \text{if } k = \ell + 1, \\ t_{ii} & \text{if } \ell = i_{r(i)} \text{ and } k = i, \\ 0 & \text{other case,} \end{cases}$$

where $t_{ii}$ is the entry of the matrix $\mathbf{T}$. Thus, in this case the matrix $\widetilde{\mathbf{T}}_{ij}$ looks like

$$\widetilde{\mathbf{T}}_{ii} = \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ t_{ii} & 0 & \dots & 0 \end{pmatrix}.$$

When $i \neq j$, let $\widetilde{\mathbf{T}}_{ij}$ denote an $r(i) \times r(j)$-dimensional matrix where its entries are defined as

$$\widetilde{t}_{\ell,k} = \begin{cases} t_{ij} & \text{if } \ell = i_{r(i)} \text{ and } k = j_1, \\ 0 & \text{other case,} \end{cases}$$

where $t_{ij}$ is the entry of $\mathbf{T}$.

Then the matrix $\widetilde{\mathbf{T}}_{ij}$ looks like

$$\widetilde{\mathbf{T}}_{ij} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ t_{ij} & 0 & \dots & 0 \end{pmatrix}.$$

For $i \in \mathcal{E}_+$ and $k \in \mathcal{E}_0$, let $\widetilde{\mathbf{T}}_{ik}$ denote a column vector of size $r(i)$ and let $\widetilde{\mathbf{T}}_{ki}$ denote a row vector of size $r(i)$, which are defined next. In the case of the column vector, the entries are defined by

$$\widetilde{t}_{\ell l} = \begin{cases} t_{ik} & \text{if } \ell = i_{r(i)} \text{ and } l = 1, \\ 0 & \text{other case,} \end{cases}$$

where $t_{ik}$ is the entry of the matrix $\mathbf{T}$. Then,

$$\widetilde{\mathbf{T}}_{ik} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ t_{ik} \end{pmatrix}.$$

Now, in the case of the row vector, its entries are defined by

$$\widetilde{t}_{\ell l} = \begin{cases} t_{ki} & \text{if } \ell = 1 \text{ and } l = i_1, \\ 0 & \text{other case,} \end{cases}$$

where $t_{kj}$ is the entry of the matrix $\mathbf{T}$. This is

$$\widetilde{\mathbf{T}}_{ki} = (t_{ki}, 0, \ldots, 0)\,.$$

For the case where $k, l \in \mathcal{E}_0$, let $\widetilde{\mathbf{T}}_{kl} = t_{kl}$. This is, $\widetilde{\mathbf{T}}_{kl}$ denotes the value $t_{kl}$, which is the entry of the matrix $\mathbf{T}$.

Finally, the constructed block–partitioned matrix looks like

$$\widetilde{\mathbf{T}} = \begin{pmatrix} \widetilde{\mathbf{T}}_{11} & \widetilde{\mathbf{T}}_{12} & \ldots & \widetilde{\mathbf{T}}_{1p} \\ \widetilde{\mathbf{T}}_{21} & \widetilde{\mathbf{T}}_{22} & \ldots & \widetilde{\mathbf{T}}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{\mathbf{T}}_{p1} & \widetilde{\mathbf{T}}_{p2} & \ldots & \widetilde{\mathbf{T}}_{pp} \end{pmatrix},$$

and it corresponds to the sub-transition probability matrix of $\{X^\star_m\}_{m\in\mathbb{N}}$.

Now, let us denote by $\widetilde{\mathbf{r}}$ a vector of rewards of size $\sum_{i=1}^{p} r(i) + |\mathcal{E}_0|$ where every entry is the reward of one of the states in $\mathcal{E}_0 \cup \left( \bigcup_{i\in\mathcal{E}_+} \mathbf{h}_i \right)$.

States in $\bigcup_{i\in\mathrm{E}_+} \mathbf{h}_i$ have reward equal to $1$ and states in $\mathcal{E}_0$ have reward equal to $0$.

Consider $\widetilde{\tau} = \min\{m \geq 0 : X^\star_m = p+1\}$. Then we can write

$$Y = \sum_{m=0}^{\widetilde{\tau}-1} \widetilde{r}\,(X^\star_m)\,. \tag{5.1}$$

By Theorem 5.1, a DPH-representation for $Y$ is given by $(\boldsymbol{\alpha}^\star, \mathbf{T}^\star)$, where

$$\boldsymbol{\pi}^\star = \widetilde{\boldsymbol{\pi}}_+ + \widetilde{\boldsymbol{\pi}}_0 \left(\mathbf{I} - \widetilde{\mathbf{T}}_{00}\right)^{-1} \widetilde{\mathbf{T}}_{0+}, \tag{5.2}$$

$$\mathbf{T}^\star = \widetilde{\mathbf{T}}_{++} + \widetilde{\mathbf{T}}_{+0} \left(\mathbf{I} - \widetilde{\mathbf{T}}_{00}\right)^{-1} \widetilde{\mathbf{T}}_{0+}, \tag{5.3}$$

and where $\widetilde{\boldsymbol{\pi}}_+$ and $\widetilde{\boldsymbol{\pi}}_0$ are the row vectors obtained by a partition of the vector $\widetilde{\boldsymbol{\pi}}$ according to the states with $1$ or $0$ rewards, so that

$$\widetilde{\boldsymbol{\pi}} = (\widetilde{\boldsymbol{\pi}}_+, \widetilde{\boldsymbol{\pi}}_0)$$

and $\widetilde{\mathbf{T}}_{++}, \widetilde{\mathbf{T}}_{+0}, \widetilde{\mathbf{T}}_{0+}$ and $\widetilde{\mathbf{T}}_{00}$ are also obtained from a partition based on the $0-1$ rewards but on the matrix $\widetilde{\mathbf{T}}$ (see Theorem 5.1).

$$\widetilde{\mathbf{T}} = \begin{pmatrix} \widetilde{\mathbf{T}}_{++} & \widetilde{\mathbf{T}}_{+0} \\ \widetilde{\mathbf{T}}_{0+} & \widetilde{\mathbf{T}}_{00} \end{pmatrix}. \tag{5.4}$$

We are now ready to introduce the next theorem.

**Theorem 5.2** *Let $\tau \sim \mathrm{DPH}_p(\boldsymbol{\pi}, \mathbf{T})$ and let $\mathbf{r} = (r(1), ..., r(p)) \in \mathbb{N}_0^p$ be the vector of rewards. Then, the distribution of*

$$Y = \sum_{m=0}^{\tau-1} r(X_m)$$

*is a mixture of a zero point and a discrete phase-type distribution with representation given by $\mathrm{DPH}_q(\boldsymbol{\pi}^\star, \mathbf{T}^\star)$, where $q = \sum_{i=1}^{p} r(i)$, and $\boldsymbol{\pi}^\star$ and $\mathbf{T}^\star$ are given in Equations* (5.2) *and* (5.3)*, respectively. The zero point is of size $1 - \boldsymbol{\pi}^\star \mathbf{e}$.*

## 5.2 The class of MDPH*

### 5.2.1 Definition

For every $j = 1, 2, \ldots, n$, let us consider $\mathbf{r}_j = (r_j(1), r_j(2), \ldots, r_j(p))^\top$, a column vector of rewards taking values in $\mathbb{N}_0^p$, and the random variable

$$Y_j = \sum_{m=0}^{\tau-1} r_j(X_m), \tag{5.5}$$

where $\tau \sim \mathrm{DPH}_p(\boldsymbol{\pi}, \mathbf{T})$.

The random vector given by

$$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$$

is said to have a Multivariate discrete phase-type distribution with representation $(\boldsymbol{\pi}, \mathbf{T}; \mathbf{R})$, where

$$\mathbf{R} = \begin{pmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n \end{pmatrix}$$

and is called the "matrix of rewards". This is denoted by $\mathbf{Y} \sim \mathrm{MDPH}_p^* (\boldsymbol{\pi}, \mathbf{T}; \mathbf{R})$.

### 5.2.2 Marginal distributions

**Corollary 5.3** *Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) \sim \mathrm{MDPH}_p^* (\boldsymbol{\pi}, \mathbf{T}; \mathbf{R})$ and let $\mathbf{w}$ be a vector taking values in $\mathbb{N}_0^n$. Then, a discrete phase-type representation for $\langle \mathbf{Y}, \mathbf{w} \rangle$ is obtained from Theorem 5.2.*

**Proof.** Let $\mathbf{w} = (w_1, w_2, \ldots, w_n) \in \mathbb{N}_0^n$. Note that

$$\langle \mathbf{Y}, \mathbf{w} \rangle = \sum_{j=1}^{n} Y_j w_j = \sum_{t=0}^{\tau-1} \sum_{j=1}^{n} r_j (X_t) w_j,$$

where $\tau \sim \mathrm{DPH}_p(\boldsymbol{\pi}, \mathbf{T})$.

Consider the matrix given by

$$\mathbf{R}_{\mathbf{w}} = \begin{pmatrix} \mathbf{r}_1 w_1 & \mathbf{r}_2 w_2 & \cdots & \mathbf{r}_n w_n \end{pmatrix}$$

and from there let us obtain the column vector

$$\mathbf{R}_{\mathbf{w}} \mathbf{e} = (R_w(1), R_w(2), \ldots, R_w(p))^\top,$$

where $\mathbf{e}$ is the column vector of ones of appropriate dimension and $R_w(i) = \sum_{j=1}^{n} r_j(i) w_j$ for every $i = 1, \ldots, p$.

Therefore, we can write

$$\langle \mathbf{Y}, \mathbf{w} \rangle = \sum_{t=0}^{\tau-1} R_w (X_t)$$

and consequently, by Theorem 5.2, we can get a discrete phase-type representation for $\langle \mathbf{Y}, \mathbf{w} \rangle$. □

### 5.2.3 Joint probability-generating function

**Theorem 5.4** *Consider* $\tau \sim \mathrm{DPH}_p(\boldsymbol{\pi}, \mathbf{T})$ *with* $\mathbb{P}(\tau = 0) = \pi_{p+1}$. *Let* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) \sim \mathrm{MDPH}_p^*(\boldsymbol{\pi}, \mathbf{T}; \mathbf{R})$. *Then, the joint probability-generating function*

$$\mathbb{E}\left(\theta_1^{Y_1} \theta_2^{Y_2} \cdots \theta_n^{Y_n}\right)$$

*exists for any* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n) \in \mathbb{R}^n$ *with* $\max\{|\theta_1|, |\theta_2|, \ldots, |\theta_n|\} \leq 1$, *and it is given by*

$$\mathbb{E}\left(\theta_1^{Y_1} \theta_2^{Y_2} \cdots \theta_n^{Y_n}\right) = \pi_{p+1} + \boldsymbol{\pi} \boldsymbol{\Delta}(\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T} \boldsymbol{\Delta}(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}, \tag{5.6}$$

*where* $\mathbf{t} = \mathbf{e} - \mathbf{T}\mathbf{e}$ *and* $\boldsymbol{\Delta}(\boldsymbol{\theta}^{\mathbf{r}})$ *denotes the diagonal matrix formed with the entries of the vector*

$$\boldsymbol{\theta}^{\mathbf{r}} = \left(\theta^{\mathbf{r}(1)}, \theta^{\mathbf{r}(2)}, \ldots, \theta^{\mathbf{r}(p)}\right),$$

*and*

$$\theta^{\mathbf{r}(i)} = \prod_{j=1}^{n} \theta_j^{r_j(i)}$$

*for every* $i = 1, \ldots, p$.

**Proof.** Assume that $X_0 = i$. Then, by Equation (5.5), $Y_j$ can be written as

$$Y_j = r_j(i) + Y_{j,1},$$

where

$$Y_{j,1} = \sum_{t=1}^{\tau-1} r_j(X_t)$$

for every $j = 1, 2, \ldots, n$, and $\tau \sim \text{DPH}_p(\boldsymbol{\pi}, \mathbf{T})$.

Next, we are going to calculate the conditional joint probability-generating function.

$$\mathbb{E}\left(\theta_1^{Y_1}\theta_2^{Y_2}\cdots\theta_n^{Y_n}\middle|X_0 = i\right) = \left(\theta_1^{r_1(i)}\theta_2^{r_2(i)}\cdots\theta_n^{r_n(i)}\right)\mathbb{E}\left(\theta_1^{Y_{1,1}}\theta_2^{Y_{2,1}}\cdots\theta_n^{Y_{n,1}}\middle|X_0 = i\right). \quad (5.7)$$

Now, we focus on the conditional joint expectation on the right side of Equation (5.7).

$$\begin{aligned}
& \mathbb{E}\left(\theta_1^{Y_{1,1}}\theta_2^{Y_{2,1}}\cdots\theta_n^{Y_{n,1}}\middle|X_0 = i\right) \\
= \; & \mathbb{E}_i\left(\theta_1^{Y_{1,1}}\theta_2^{Y_{2,1}}\cdots\theta_n^{Y_{n,1}}\right) \\
= \; & \sum_{\xi=1}^{p+1}\mathbb{E}_i\left(\mathbf{1}_{\{X_1=\xi\}}\theta_1^{Y_{1,1}}\theta_2^{Y_{2,1}}\cdots\theta_n^{Y_{n,1}}\right) \\
= \; & \mathbb{E}_i\left(\mathbf{1}_{\{X_1=p+1\}}\theta_1^{Y_{1,1}}\theta_2^{Y_{2,1}}\cdots\theta_n^{Y_{n,1}}\middle|X_1 = p+1\right)\mathbb{P}_i\left(X_1 = p+1\right) \\
+ \; & \sum_{\xi=1}^{p}\mathbb{E}_i\left(\mathbf{1}_{\{X_1=\xi\}}\theta_1^{Y_{1,1}}\theta_2^{Y_{2,1}}\cdots\theta_n^{Y_{n,1}}\middle|X_1 = \xi\right)\mathbb{P}_i\left(X_1 = \xi\right) \qquad (5.8) \\
= \; & t_{i,p+1} + \sum_{\xi=1}^{p}\mathbb{E}\left(\theta_1^{Y_1}\theta_2^{Y_2}\cdots\theta_n^{Y_n}\middle|X_0 = \xi\right)t_{i,\xi} \quad \text{due to the homogeneity,} \\
= \; & t_{i,p+1} + \sum_{\xi=1}^{p}\mathbb{E}_\xi\left(\theta_1^{Y_1}\theta_2^{Y_2}\cdots\theta_n^{Y_n}\right)t_{i,\xi}. \qquad (5.9)
\end{aligned}$$

Let

$$\begin{aligned}
\theta^{\mathbf{Y}} &= \theta_1^{Y_1}\theta_2^{Y_2}\cdots\theta_n^{Y_n}, \\
\theta^{\mathbf{r}(\xi)} &= \theta_1^{r_1(\xi)}\theta_2^{r_2(\xi)}\cdots\theta_n^{r_n(\xi)},
\end{aligned}$$

$$\theta^{-\mathbf{r}(\xi)} = \theta_1^{-r_1(\xi)}\theta_2^{-r_2(\xi)}\cdots\theta_n^{-r_n(\xi)}.$$

Then, by combining Equations (5.7) and (5.8), we obtain an expression for the next expectation

$$\mathbb{E}_i\left(\theta^{\mathbf{Y}}\right) = \theta^{\mathbf{r}(i)}\left(t_{i,p+1} + \sum_{\xi=1}^{p}\mathbb{E}_\xi\left(\theta^{\mathbf{Y}}\right)t_{i,\xi}\right),$$

which is equivalent to

$$\theta^{-\mathbf{r}(i)}\mathbb{E}_i\left(\theta^{\mathbf{Y}}\right) = t_{i,p+1} + \sum_{\xi=1}^{p}\mathbb{E}_\xi\left(\theta^{\mathbf{Y}}\right)t_{i,\xi}, \tag{5.10}$$

where $i \in \{1, 2, \ldots, p\}$.

The system of equations formed from Equation (5.10) can be expressed in the following matrix form

$$\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{-\mathbf{r}}\right)\mathbf{E}(\boldsymbol{\theta}^{\mathbf{Y}}) = \mathbf{t} + \mathbf{T}\mathbf{E}(\boldsymbol{\theta}^{\mathbf{Y}}), \tag{5.11}$$

where

$$\boldsymbol{\theta}^{-\mathbf{r}} = \left(\theta^{-\mathbf{r}(1)}, \theta^{-\mathbf{r}(2)}, \ldots, \theta^{-\mathbf{r}(p)}\right),$$

and

$$\mathbf{E}(\boldsymbol{\theta}^{\mathbf{Y}}) = \left(\mathbb{E}_1\left(\theta^{\mathbf{Y}}\right), \mathbb{E}_2\left(\theta^{\mathbf{Y}}\right), \ldots, \mathbb{E}_p\left(\theta^{\mathbf{Y}}\right)\right)^\top.$$

Now, the system in Equation (5.11) can also be written as

$$\Big(\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{-\mathbf{r}}\right) - \mathbf{T}\Big)\mathbf{E}(\boldsymbol{\theta}^{\mathbf{Y}}) = \mathbf{t}. \tag{5.12}$$

Since $\max\{|\theta_1|, |\theta_2|, \ldots, |\theta_n|\} \le 1$, then the matrix $\Big(\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{-\mathbf{r}}\right) - \mathbf{T}\Big)$ is nonsingular and consequently the system in Equation (5.12) is equal to

$$\mathbf{E}(\boldsymbol{\theta}^{\mathbf{Y}}) = \Big(\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{-\mathbf{r}}\right) - \mathbf{T}\Big)^{-1}\mathbf{t},$$

or equivalently to

$$\mathbf{E}(\boldsymbol{\theta}^{\mathbf{Y}}) = \boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\Big(\mathbf{I} - \mathbf{T}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\Big)^{-1}\mathbf{t}.$$

Finally, the joint probability-generating function is given by

$$\mathbb{E}\left(\theta_1^{Y_1}\theta_2^{Y_2}\cdots\theta_n^{Y_n}\right) = \pi_{p+1} + \boldsymbol{\pi}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\Big(\mathbf{I} - \mathbf{T}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\Big)^{-1}\mathbf{t}.$$

□

**Note:** In order to justify that the matrix $(\mathbf{I} - \mathbf{T}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right))$ is nonsingular, we are going to use the following theorem which can be found in [Hog07].

**Theorem 5.5** *The following statements are equivalent.*

*(1)* $\rho(\mathbf{T}) < 1$.

*(2)* $\mathbf{I} - \mathbf{T}$ *is nonsingular.*

*(3) There exists a positive vector* $\mathbf{u} \in \mathbb{R}^p$ *such that* $\mathbf{Tu} < \mathbf{u}$.

From point (3), there exists the vector $\mathbf{u} = (u_1, u_2, \ldots, u_p)^\top$ such that $\mathbf{Tu} < \mathbf{u}$. Then

$$\begin{aligned}
\mathbf{T}\boldsymbol{\Delta}(\boldsymbol{\theta}^{\mathbf{r}})\,\mathbf{u} &= \mathbf{T} \cdot \left(u_1\theta^{\mathbf{r}(1)}, u_2\theta^{\mathbf{r}(2)}, \ldots, u_p\theta^{\mathbf{r}(p)}\right)^\top \\
&\leq \mathbf{T} \cdot (u_1, u_2, \ldots, u_p)^\top \theta^* \quad \text{where} \quad \theta^* = \max_{1 \leq i \leq p} \left\{|\theta|^{\mathbf{r}(i)}\right\} \\
&\leq \mathbf{T} \cdot (u_1, u_2, \ldots, u_p)^\top \quad \text{since} \quad \theta^* \leq 1 \\
&= \mathbf{Tu} \\
&< \mathbf{u} \quad \text{by assumption.}
\end{aligned}$$

Therefore, by Theorem 5.5, the matrix $(\mathbf{I} - \mathbf{T}\boldsymbol{\Delta}(\boldsymbol{\theta}^{\mathbf{r}}))$ is nonsingular.

### 5.2.4 Moments

**Corollary 5.6 (Moment-generating function.)** *Let* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) \sim$ MDPH$^*(\boldsymbol{\pi}, \mathbf{T}; \mathbf{R})$. *Then, the joint moment-generating function of* $\mathbf{Y}$ *is given by*

$$\mathbb{E}\left(\exp\left(\sum_{i=1}^{n} u_i Y_i\right)\right) = \boldsymbol{\pi}\boldsymbol{\Delta}(\mathbf{e}^{\mathbf{u}\cdot\mathbf{r}})\left(\mathbf{I} - \mathbf{T}\boldsymbol{\Delta}(\mathbf{e}^{\mathbf{u}\cdot\mathbf{r}})\right)^{-1}\mathbf{t}, \tag{5.13}$$

*where* $\mathbf{u} = (u_1, \ldots, u_n)$, $u_i \leq 0$, *for all* $i = 1, \ldots, n$, *and* $\mathbf{e}^{\mathbf{u}\cdot\mathbf{r}}$ *denotes the vector*

$$\left(\exp\left(\sum_{i=1}^{n} u_i r_i(1)\right), \exp\left(\sum_{i=1}^{n} u_i r_i(2)\right), \ldots, \exp\left(\sum_{i=1}^{n} u_i r_i(p)\right)\right).$$

**Proof.** It follows directly from Equation (5.6) and from the change of variable $\theta_i = e^{u_i}$, where $u_i \leq 0$, for all $i = 1, 2, \ldots, n$. □

In the next, we are going to derive a closed-form formula for general moments of a marginal of a MDPH*-distributed random vector.

First, let us write formula (5.13) as follows

$$\mathbb{E}\left(\exp\left(\sum_{i=1}^{n} u_i Y_i\right)\right) = \boldsymbol{\pi}\Big(\boldsymbol{\Delta}\left(\mathbf{e}^{-\mathbf{u}\cdot\mathbf{r}}\right) - \mathbf{T}\Big)^{-1}\mathbf{t}. \tag{5.14}$$

Consequently, for every fixed $j = 1, \ldots, n$, the $m$-th derivative with respect to the variable $u_j$ of the moment-generating function in Equation (5.14) is given by

$$\frac{d^m \mathbb{E}\left(\exp\left(\sum_{i=1}^{n} u_i Y_i\right)\right)}{du_j^m} = \boldsymbol{\pi}\left[\frac{d^m \Big(\boldsymbol{\Delta}\left(\mathbf{e}^{-\mathbf{u}\cdot\mathbf{r}}\right) - \mathbf{T}\Big)^{-1}}{du_j^m}\right]\mathbf{t}, \tag{5.15}$$

see Equation (A.1).

From the formula of the $m$-th derivative of an inverse matrix (see Equation (A.2)), we obtain

$$\frac{d^m \left(\boldsymbol{\Delta}\left(\mathbf{e}^{-\mathbf{u}\cdot\mathbf{r}}\right) - \mathbf{T}\right)^{-1}}{du_j^m} =$$

$$= m! \sum_{s=1}^{m} (-1)^s \sum_{\substack{1 \le n_1, n_2, \ldots, n_s \le m \\ n_1 + \cdots + n_s = m}} \Big(\boldsymbol{\Delta}\left(\mathbf{e}^{-\mathbf{u}\cdot\mathbf{r}}\right) - \mathbf{T}\Big)^{-1} \prod_{r=1}^{s} \left(\frac{1}{n_r!}\left(\frac{d^{n_r}\mathbf{L}}{du_j^{n_r}}\right)\Big(\boldsymbol{\Delta}\left(\mathbf{e}^{-\mathbf{u}\cdot\mathbf{r}}\right) - \mathbf{T}\Big)^{-1}\right),$$

where

$$\frac{d^{n_r}\mathbf{L}}{du_j^{n_r}} = \frac{d^{n_r}\Big(\boldsymbol{\Delta}\left(\mathbf{e}^{-\mathbf{u}\cdot\mathbf{r}}\right) - \mathbf{T}\Big)}{du_j^{n_r}} = \boldsymbol{\Delta}^{n_r}\left(-\mathbf{r}_j\right)\boldsymbol{\Delta}\left(\mathbf{e}^{-\mathbf{u}\cdot\mathbf{r}}\right).$$

Then

$$\left.\frac{d^{n_r}\mathbf{L}}{du_j^{n_r}}\right|_{u_i = 0, \forall i = 1, \ldots, n.} = \boldsymbol{\Delta}^{n_r}\left(-\mathbf{r}_j\right),$$

where $\mathbf{r}_j = (r_j(1), \ldots, r_j(p))^\top$ and

$$\left.\Big(\boldsymbol{\Delta}\left(\mathbf{e}^{-\mathbf{u}\cdot\mathbf{r}}\right) - \mathbf{T}\Big)^{-1}\right|_{u_i = 0, \forall i = 1, \ldots, n.} = \left(\mathbf{I} - \mathbf{T}\right)^{-1}.$$

Therefore,

$$\left.\frac{d^m \left(\boldsymbol{\Delta}\left(\mathbf{e}^{-\mathbf{u}\cdot\mathbf{r}}\right) - \mathbf{T}\right)^{-1}}{du_j^m}\right|_{u_i = 0, \forall i = 1, \ldots, n.}$$

$$= \quad m! \sum_{s=1}^{m} (-1)^s \sum_{\substack{1 \le n_1, n_2, \ldots, n_s \le m \\ n_1 + \cdots + n_s = m}} (\mathbf{I} - \mathbf{T})^{-1} \prod_{r=1}^{s} \left( \frac{1}{n_r!} \mathbf{\Delta}^{n_r} (-\mathbf{r}_j) (\mathbf{I} - \mathbf{T})^{-1} \right).$$

Let us denote

$$\mathbf{U} = (\mathbf{I} - \mathbf{T})^{-1}.$$

Consequently, we have

$$\begin{aligned}
\mathbb{E}\left(Y_j^m\right) &= m!\boldsymbol{\pi} \sum_{s=1}^{m} (-1)^s \sum_{\substack{1 \le n_1, n_2, \ldots, n_s \le m \\ n_1 + \cdots + n_s = m}} \prod_{r=1}^{s} \left( \frac{1}{n_r!} \mathbf{U} \mathbf{\Delta}^{n_r} (-\mathbf{r}_j) \right) \mathbf{e} \\
&= m!\boldsymbol{\pi} \sum_{s=1}^{m} (-1)^s \sum_{\substack{1 \le n_1, n_2, \ldots, n_s \le m \\ n_1 + \cdots + n_s = m}} \prod_{r=1}^{s} \left( \frac{(-1)^{n_r}}{n_r!} \mathbf{U} \mathbf{\Delta} \left( \mathbf{r}_j^{n_r} \right) \right) \mathbf{e} \\
&= (-1)^m m!\boldsymbol{\pi} \sum_{s=1}^{m} (-1)^s \sum_{\substack{1 \le n_1, n_2, \ldots, n_s \le m \\ n_1 + \cdots + n_s = m}} \prod_{r=1}^{s} \left( \mathbf{U} \mathbf{\Delta} \left( \frac{\mathbf{r}_j^{n_r}}{n_r!} \right) \right) \mathbf{e},
\end{aligned}$$

where $\frac{\mathbf{r}_j^{n_r}}{n_r!}$ means that every entry of the vector $\mathbf{r}_j^{n_r}$ is power $n_r$ and divided by $n_r!$.

Now, let us introduce the set of permutations given by

$$\mathcal{P}_m = \left\{ (\sigma(1), \ldots, \sigma(m)) : 0 \le \sigma(s) \le m, \sum_{l=1}^{m} \sigma(l) = m \right\}.$$

Let us take one permutation given by

$$\boldsymbol{\sigma}_\ell = (\sigma_\ell(1), \ldots, \sigma_\ell(m)) \in \mathcal{P}_m.$$

We define

$$(\mathbf{U} \mathbf{\Delta} (\mathbf{r}_j))_{\sigma_\ell(s)} = \begin{cases} -\mathbf{U} \mathbf{\Delta} \left( \frac{\mathbf{r}_j^{\sigma_\ell(s)}}{\sigma_\ell(s)!} \right) & \text{for } \sigma_\ell(s) \neq 0, \\ \mathbf{I} & \text{for } \sigma_\ell(s) = 0, \end{cases} \tag{5.16}$$

where $\mathbf{I}$ denotes the identity matrix of appropriate dimension.

Thus, we can write

$$\mathbb{E}\left(Y_j^m\right) = (-1)^m m!\boldsymbol{\pi} \sum_{\ell=1}^{m!} \left( \prod_{s=1}^{m} (\mathbf{U} \mathbf{\Delta} (\mathbf{r}_j))_{\sigma_\ell(s)} \right) \mathbf{e},$$

and finally, we conclude with the following proposition.

**Proposition 5.7 (The m-th moment.)** *Let* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) \sim \mathrm{MDPH}^* (\boldsymbol{\pi}, \mathbf{T}; \mathbf{R})$. *Then, for every fixed* $j = 1, \ldots, n$, *it holds*

$$\mathbb{E}\left(Y_j^m\right) = (-1)^m m! \boldsymbol{\pi} \sum_{\ell=1}^{m!} \left( \prod_{s=1}^{m} \left(\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\right)_{\sigma_\ell(s)} \right) \mathbf{e}, \tag{5.17}$$

*where* $\mathbf{r}_j = (r_j(1), r_j(2), \ldots, r_j(p))^\top$, $\mathbf{U} = (\mathbf{I} - \mathbf{T})^{-1}$ *and* $\left(\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\right)_{\sigma_\ell(s)}$ *is defined in Equation* (5.16).

In the following, by using Equation (5.17) we show formulas for the first moments.

**First moment.**

$$\mathbb{E}\left(Y_j\right) = \boldsymbol{\pi}\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\mathbf{e}.$$

**Second moment.**

$$\mathbb{E}\left(Y_j^2\right) = \boldsymbol{\pi}\left\{-\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j^2\right) + 2\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\right\}\mathbf{e}.$$

**Cross moments in the bivariate case.**

$$\begin{aligned}\mathbb{E}\left(Y_j Y_k\right) &= \boldsymbol{\pi}\Big\{\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_k\right) + \mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_k\right)\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right) \\ &\quad - \mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\boldsymbol{\Delta}\left(\mathbf{r}_k\right)\Big\}\mathbf{e}.\end{aligned}$$

**Covariance.**

$$\begin{aligned}\mathrm{Cov}\left(Y_j, Y_k\right) &= \boldsymbol{\pi}\Big\{\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_k\right) + \mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_k\right)\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right) \\ &\quad - \mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\boldsymbol{\Delta}\left(\mathbf{r}_k\right) - \mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_j\right)\mathbf{e}\boldsymbol{\pi}\mathbf{U}\boldsymbol{\Delta}\left(\mathbf{r}_k\right)\Big\}\mathbf{e}.\end{aligned}$$

### 5.2.5 Closure properties of the class

**Proposition 5.8**

*(a) Consider* $n$ *independent and DPH-distributed random variables given by* $Y_j \sim \mathrm{DPH}_p\left(\boldsymbol{\pi}_j, \mathbf{T}_j\right)$, $j = 1, \ldots, n$. *Then, the random vector* $(Y_1, Y_2, \ldots, Y_n)$ *is* MDPH***-*distributed.*

*(b) Assume that* $(Y_1, Y_2, \ldots, Y_n)$ *is* $\mathrm{MDPH}_p^*$*-distributed. Then, every marginal* $Y_j$ *is* DPH*-distributed,* $j = 1, \ldots, n$.

*(c) Let* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) \sim \text{MDPH}_p^* (\boldsymbol{\pi}, \mathbf{T}; \mathbf{R})$ *, where*

$$\mathbf{R} = \begin{pmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n \end{pmatrix}$$

*and*

$$\mathbf{r}_j = (r_j(1), r_j(2), \ldots, r_j(p))^\top, j = 1, \ldots, n.$$

*Consider an* $\ell \times n$*-dimensional matrix* $\mathbf{A} = \{a_{l,j}\}$ *having non-negative entire entries. Then, the random vector* $\mathbf{A}\mathbf{Y}^\top$ *is* MDPH*-*distributed with representation given by* $(\boldsymbol{\pi}, \mathbf{T}; \mathbf{R}^*)$ *, where*

$$\begin{aligned} \mathbf{R}^* &= \begin{pmatrix} \mathbf{r}_1^* & \mathbf{r}_2^* & \cdots & \mathbf{r}_\ell^* \end{pmatrix}, \\ \mathbf{r}_k^* &= (r_k^*(1), r_k^*(2), \ldots, r_k^*(p))^\top, \quad k = 1, \ldots, \ell, \end{aligned}$$

*and* $r_k^*(i) = \sum_{j=1}^n a_{k,j} r_j(i), i = 1, \ldots, p.$

**Proof.** *(a)* It is enough to prove this statement for the case $n = 2$. Let $Y_1 \sim \text{MDPH}_p^* (\boldsymbol{\pi}_1, \mathbf{T}_1)$ and $Y_2 \sim \text{MDPH}_q^* (\boldsymbol{\pi}_2, \mathbf{T}_2)$ be independent. Consider $(\boldsymbol{\gamma}, \mathbf{G}; \mathbf{R}_\mathbf{G})$ , where

$$\boldsymbol{\gamma} = (\boldsymbol{\pi}_1, \mathbf{0}), \quad \mathbf{G} = \begin{pmatrix} \mathbf{T}_1 & \mathbf{t}_1\boldsymbol{\pi}_2 \\ \mathbf{0} & \mathbf{T}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{R}_\mathbf{G} = \begin{pmatrix} \mathbf{e}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{e}_q \end{pmatrix}.$$

Denote

$$\mathbf{g} = \begin{pmatrix} \mathbf{0} \\ \mathbf{t}_2 \end{pmatrix}.$$

Then, we have that

$$\begin{aligned} & \boldsymbol{\gamma}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}_\mathbf{g}}\right)\left(\mathbf{I} - \mathbf{G}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}_\mathbf{g}}\right)\right)^{-1}\mathbf{g} \\ = \; & \text{the product of the two univariate probability-generating function of } Y_1 \text{ and } Y_2. \end{aligned}$$

*(b)* Since $(Y_1, Y_2, \ldots, Y_n)$ is MDPH*-distributed, then every variable $Y_j, j = 1, \ldots, n$, can be written as the accumulated rewards obtained before the Markov chain gets absorbed (by definition). That is,

$$Y_j = \sum_{t=0}^{\tau-1} r_j(X_t),$$

where $\tau$ is DPH-distributed (see equation (5.5)) and for some vector of rewards

$$(r_j(1), \ldots, r_j(p))^\top.$$

Finally, Theorem 5.2 provides a DPH-representation for $Y_j$.

*(c)* The random vector $\mathbf{AY}^\top$ is given by

$$\mathbf{AY}^\top = \left( \sum_{j=1}^{n} a_{1,j} Y_j, \sum_{j=1}^{n} a_{2,j} Y_j, \ldots, \sum_{j=1}^{n} a_{\ell,j} Y_j \right)^\top . \tag{5.18}$$

In order to obtain the matrix of rewards, consider the random variable $N_i$ which represents the number of times the underlying Markov chain visits the state $i$ prior to absorption. Then, every entry in $\mathbf{AY}^\top$ can be written as follows. For every $k = 1, 2, \ldots, \ell$,

$$\begin{aligned} \sum_{j=1}^{n} a_{k,j} Y_j &= \sum_{j=1}^{n} a_{k,j} \sum_{i=1}^{p} r_j(i) N_i \\ &= \sum_{i=1}^{p} \left( \sum_{j=1}^{n} a_{k,j} r_j(i) \right) N_i. \end{aligned}$$

Therefore, the reward in state $i$ of the random variable in the $k$-th entry is given by $\sum\limits_{j=1}^{n} a_{k,j} r_j(i)$. □

**Corollary 5.9 (Finite convolutions and finite conjunctions.)**

*i. The class of* MDPH$^*$ *is closed under finite convolutions. That is, if*

$$(Y_1, Y_2, \ldots, Y_n) \sim \text{MDPH}^* \left( \boldsymbol{\pi}, \mathbf{T}; \mathbf{R} \right),$$

*then* $\sum\limits_{j=1}^{n} Y_j$ *is* DPH*-distributed.*

*ii. The class of* MDPH$^*$ *is closed under finite conjunctions. That is, if* $\mathbf{Z}_1, \ldots \mathbf{Z}_n$ *are independent and* MDPH$^*$*-distributed random vectors, then* $(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ *is also a* MDPH$^*$*-distributed random vector.*

**Proof.** $i$. By taking $\mathbf{r} = \mathbf{Re}$ as the vector of rewards, the result follows from Theorem 5.2.

$ii$.The proof is similar to the statement *(a)* in the Proposition 5.8.

□

**Theorem 5.10 (Finite mixtures.)** *The class of* MDPH$^*$ *is closed under finite mixtures.*

**Proof.** Let $\{p_1, \ldots, p_m\}$ be a probability mass function and consider $m$ MDPH*-distributed random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_m$ with representation given by $(\boldsymbol{\pi}_i, \mathbf{T}_i; \mathbf{R}_i)$, $i = 1, \ldots, m$, respectively. Let us denote by $F_i$ the distribution function of the random vector $\mathbf{X}_i$, for every $i = 1, \ldots, m$. We are going to prove that

$$\sum_{i=1}^{m} p_i F_i$$

is MDPH$^*$-distributed.

Consider the case $m = 2$. Thus, we have two MDPH*-distributed random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$. Then, consider a Markov chain with initial distribution given by

$$(p_1 \boldsymbol{\pi}_1, (1 - p_1)\boldsymbol{\pi}_2)$$

and sub-transition probability matrix

$$\begin{pmatrix} \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 \end{pmatrix}.$$

Finally, by choosing the matrix of rewards as

$$\begin{pmatrix} \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 \end{pmatrix}$$

the proof is completed.

□

As a particular case of finite mixtures of MDPH$^*$-distributions, we have the subclass of mixtures of multivariate negative binomial distributions, which is dense in the set of multivariate distributions with support in $\mathbb{N}_0 \times \cdots \times \mathbb{N}_0$ (in the sense of weak convergence of distributions). Additionally, in [ALSS84] it is explained that multivariate phase-type distributions taking values in $[0, \infty)^n, n \in \mathbb{N}$, are dense in the set of distributions with support $[0, \infty)^n$ and the same statement is applied for the discrete case.

**Corollary 5.11 (Denseness.)** *The class of* MDPH$^*$ *is dense in the set of multivariate distributions with support on* $\mathbb{N}_0 \times \cdots \times \mathbb{N}_0$.

# 5.3 Examples

## 5.3.1 Bivariate Geometric distribution of order $k$.

**Definition 5.12** Consider the sequence $\{B_i\}_{i \in \mathbb{N}}$ of i.i.d. random variables with distribution $B_i \sim$Bernoulli$(p)$. We denote $\mathbb{P}(B_i = 1) = p$, $\mathbb{P}(B_i = 0) = q$ and $q = 1 - p$.

Let $T$ denote the time until $k$ ones appear successively and let $M_0$ be the number of zeros that appears before the $k$ ones are obtained in a row. Then, the random vector $(M_{_0}, T)$ is said to be **bivariate geometric distributed of order** $k$ and its joint probability-generating function of $(M_{_0}, T)$ is given by

$$\mathbb{E}\left(\theta_1^{M_0}\theta_2^T\right) = \frac{(\theta_2 p)^k}{1 - q\theta_1\theta_2 \sum\limits_{i=0}^{k-1}(p\theta_2)^i}.$$

See [BKJ97, p. 117].

The distribution of $(M_{_0}, T)$ is a bivariate discrete phase-type with a representation given by

$$(\boldsymbol{\pi}_{k+1}, \mathbf{T}_{k+1}; \mathbf{R}_{k+1}), \tag{5.19}$$

where

$$\boldsymbol{\pi}_{k+1} = (1, 0, \ldots, 0),$$

is an $(k+1)$-dimensional row vector,

$$\mathbf{T}_{k+1} = \begin{pmatrix} 0 & q & p & 0 & 0 & \cdots & 0 \\ 0 & q & p & 0 & 0 & \cdots & 0 \\ 0 & q & 0 & p & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & q & 0 & 0 & 0 & \cdots & p \\ 0 & q & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

is an $(k+1) \times (k+1)$-dimensional matrix and

$$\begin{array}{cc} & \begin{array}{cc} M_{_0} & T \end{array} \\ \mathbf{R}_{k+1} = & \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \end{array}$$

is an $(k+1) \times 2$-dimensional matrix of rewards.

Consider a Markov chain $\{X_t\}_{t\in\mathbb{N}}$ with the initial distribution $\boldsymbol{\pi}_{k+1}$ and the sub-transition probability matrix $\mathbf{T}_{k+1}$. This Markov chain is a Bernoulli process that starts at the state $S$ (the "starting" state), and the corresponding rewards for this state are 0 for the variable $M_{_0}$ and 1 for the variable $T$. Then, the process can jump to the state 0 (with probability $q$), which refers to the event $\{B_1 = 0\}$, or it can jump to the state $1_1$

(with probability $p$), which refers to the event $\{B_1 = 1\}$. The rewards for the state 0 are 1 for both random variables $T$ and $M_0$. Now, if the process jumps to the state $1_1$, the rewards are 1 for the random variable $T$ and 0 for $M_0$. From state $1_1$, the process can jump to the state $1_2$ (with probability $p$), which refers to the event $\{B_2 = 1\}$, and it means that two successive ones were obtained. State $1_2$ is 1-rewarded for the variable $T$ and 0-rewarded for the variable $M_0$. In the case where the Markov chain jumps to the state $1_2$, then it can continue in the same way until it reaches the state $1_k$, which is equivalent to the event of getting $k$ ones in a row. As well as, notice that after the first jump, the Markov chain can go back to the state 0 (with probability $q$) where in that case it restarts over again.

In the next, we prove by induction that the proposed MDPH*-representation is in fact for the bivariate Geometric distribution of order $k$.

Let us start with the case $k = 2$. Then, the representation for $(M_0, T)$, is

$$(\boldsymbol{\pi}_3, \mathbf{T}_3; \mathbf{R}_3), \tag{5.20}$$

where

$$\boldsymbol{\pi}_3 = (1, 0, 0), \quad \mathbf{T}_3 = \begin{pmatrix} 0 & q & p \\ 0 & q & p \\ 0 & q & 0 \end{pmatrix}, \quad \mathbf{R}_3 = \begin{array}{c} \begin{matrix} M_0 & T \end{matrix} \\ \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \end{array}.$$

Thus, we form the matrix

$$\boldsymbol{\Delta}_3(\boldsymbol{\theta}^{\mathbf{r}}) = \begin{pmatrix} \theta_2 & 0 & 0 \\ 0 & \theta_1\theta_2 & 0 \\ 0 & 0 & \theta_2 \end{pmatrix}$$

and the exit vector

$$\mathbf{t}_3 = \begin{pmatrix} 0 \\ 0 \\ p \end{pmatrix}.$$

Now, we calculate directly inverse matrix $\Big(\mathbf{I} - \mathbf{T}_3\boldsymbol{\Delta}_3(\boldsymbol{\theta}^{\mathbf{r}})\Big)^{-1}$.

$$\Big(\mathbf{I} - \mathbf{T}_3\boldsymbol{\Delta}_3(\boldsymbol{\theta}^{\mathbf{r}})\Big)^{-1} = \begin{pmatrix} 1 & -\frac{q\theta_1\,\theta_2\,(p\theta_2+1)}{p{\theta_2}^2q\theta_1+q\theta_1\,\theta_2-1} & -\frac{p\theta_2}{p{\theta_2}^2q\theta_1+q\theta_1\,\theta_2-1} \\ 0 & -\left(p{\theta_2}^2q\theta_1+q\theta_1\,\theta_2-1\right)^{-1} & -\frac{p\theta_2}{p{\theta_2}^2q\theta_1+q\theta_1\,\theta_2-1} \\ 0 & -\frac{q\theta_1\,\theta_2}{p{\theta_2}^2q\theta_1+q\theta_1\,\theta_2-1} & \frac{q\theta_1\,\theta_2-1}{p{\theta_2}^2q\theta_1+q\theta_1\,\theta_2-1} \end{pmatrix}.$$

As well as, we can calculate

$$\det\Big(\mathbf{I} - \mathbf{T}_3\boldsymbol{\Delta}_3(\boldsymbol{\theta}^{\mathbf{r}})\Big) = -p{\theta_2}^2q\theta_1 - q\theta_1\,\theta_2 + 1 = 1 - q\theta_1\,\theta_2\sum_{i=0}^{1}(p\theta_2)^i.$$

Then, the matrix $\left(\mathbf{I}-\mathbf{T}_3\boldsymbol{\Delta}_3(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1}$ can be written as

$$\left(\mathbf{I}-\mathbf{T}_3\boldsymbol{\Delta}_3(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} = \frac{1}{\det\left(\mathbf{I}-\mathbf{T}_3\boldsymbol{\Delta}_3(\boldsymbol{\theta}^{\mathbf{r}})\right)}\mathbf{C}_3^{\top},$$

where

$$\mathbf{C}_3^{\top} = \begin{pmatrix} -p{\theta_2}^2 q\theta_1 - q\theta_1\,\theta_2 + 1 & q\theta_1\,\theta_2\,(p\theta_2+1) & p\theta_2 \\ 0 & 1 & p\theta_2 \\ 0 & q\theta_1\,\theta_2 & -q\theta_1\,\theta_2+1 \end{pmatrix},$$

($\mathbf{C}_3$ is the matrix of cofactors).

Now, by computing

$$\boldsymbol{\pi}_3\boldsymbol{\Delta}_3(\boldsymbol{\theta}^{\mathbf{r}})\left(\mathbf{I}-\mathbf{T}_3\boldsymbol{\Delta}_3(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1}\mathbf{t}_3 = \frac{(p\theta_2)^2}{1-q\theta_1\,\theta_2\sum\limits_{i=0}^{1}(p\theta_2)^i},$$

we can see that we only need the value of the entry $c_{31}$ of the matrix $\mathbf{C}_3$.

**Induction hypothesis.**

For $k-1\in\mathbb{N}$, assume the following two points.

1. $\det\left(\mathbf{I}-\mathbf{T}_k\boldsymbol{\Delta}_k(\boldsymbol{\theta}^{\mathbf{r}})\right) = 1-q\theta_1\,\theta_2\sum\limits_{i=0}^{k-2}(p\theta_2)^i$.
2. $c_{k,1}=c_{k,2}=(p\theta_2)^{k-2}$, where $c_{k,1}$ and $c_{k,2}$ are entries of the matrix $\mathbf{C}_k$, which is the matrix of cofactors of the matrix $(\mathbf{I}-\mathbf{T}_k\boldsymbol{\Delta}_k(\boldsymbol{\theta}^{\mathbf{r}}))$.

We are going to prove that the last two points (1. and 2.) also apply for $k\in\mathbb{N}$.

First we calculate the inverse matrix

$$\left(\mathbf{I}-\mathbf{T}_{k+1}\boldsymbol{\Delta}_{k+1}(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1}$$

and for that we write the matrix as follows

$$\mathbf{I}-\mathbf{T}_{k+1}\boldsymbol{\Delta}_{k+1}(\boldsymbol{\theta}^{\mathbf{r}}) = \begin{pmatrix} \left(\mathbf{I}-\mathbf{T}_k\boldsymbol{\Delta}_k(\boldsymbol{\theta}^{\mathbf{r}})\right) & \mathbf{b} \\ \mathbf{c} & 1 \end{pmatrix},$$

where

$$\mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -p\theta_2 \end{pmatrix} \text{ and } \mathbf{c} = (0, -q\theta_1\theta_2, 0, \ldots, 0, 0)\,,$$

$\mathbf{b}$ is an $k$-dimensional column vector and $\mathbf{c}$ is an $k$-dimensional row vector.

Let us denote by $\mathbf{A}_k$ the matrix $\Big(\mathbf{I} - \mathbf{T}_k\boldsymbol{\Delta}_k(\boldsymbol{\theta}^{\mathbf{r}})\Big)$. Thus,

$$\begin{aligned}
&\Big(\mathbf{I} - \mathbf{T}_{k+1}\boldsymbol{\Delta}_{k+1}(\boldsymbol{\theta}^{\mathbf{r}})\Big)^{-1} \\
&= \begin{pmatrix} \mathbf{A}_k^{-1} + \mathbf{A}_k^{-1}\mathbf{b}\left(1 - \mathbf{c}\mathbf{A}_k^{-1}\mathbf{b}\right)^{-1}\mathbf{c}\mathbf{A}_k^{-1} & -\mathbf{A}_k^{-1}\mathbf{b}\left(1 - \mathbf{c}\mathbf{A}_k^{-1}\mathbf{b}\right)^{-1} \\ -\left(1 - \mathbf{c}\mathbf{A}_k^{-1}\mathbf{b}\right)\mathbf{c}\mathbf{A}_k^{-1} & \left(1 - c\mathbf{A}_k^{-1}\mathbf{b}\right)^{-1} \end{pmatrix}.
\end{aligned}$$

Now, observe that

$$-\mathbf{A}_k^{-1}\mathbf{b} = \frac{-1}{\det(\mathbf{A}_k)}\mathbf{C}_k^{\top}\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -p\theta_2 \end{pmatrix} = \frac{1}{\det(\mathbf{A}_k)}\begin{pmatrix} c_{k,1}p\theta_2 \\ c_{k,2}p\theta_2 \\ \vdots \\ c_{k,k-1}p\theta_2 \\ c_{k,k}p\theta_2 \end{pmatrix}. \tag{5.21}$$

By the induction hypothesis, the first entry of the vector in (5.21) is given by

$$\left(-\mathbf{A}_k^{-1}\mathbf{b}\right)_1 = \frac{(p\theta_2)^{k-1}}{\det(\mathbf{A}_k)}. \tag{5.22}$$

As well as, we apply the induction hypothesis for the next calculation

$$\mathbf{c}\left(-\mathbf{A}_k^{-1}\mathbf{b}\right) = -q\theta_1\theta_2\left(\frac{c_{k,2}p\theta_2}{\det(\mathbf{A}_k)}\right) = \frac{-q\theta_1\theta_2(p\theta_2)^{k-1}}{\det(\mathbf{A}_k)}\,,$$

where we use that $c_{k,2} = (p\theta_2)^{k-2}$. Hence, we get

$$\left(1 - \mathbf{c}\mathbf{A}_k^{-1}\mathbf{b}\right)^{-1} = \frac{\det(\mathbf{A}_k)}{\det(\mathbf{A}_k) - q\theta_1\theta_2(p\theta_2)^{k-1}}.$$

Finally,

$$-\mathbf{A}_k^{-1}\mathbf{b}\left(1 - \mathbf{c}\mathbf{A}_k^{-1}\mathbf{b}\right)^{-1} = \frac{1}{\det(\mathbf{A}_k) - q\theta_1\theta_2(p\theta_2)^{k-1}}\begin{pmatrix} c_{k,1}p\theta_2 \\ c_{k,2}p\theta_2 \\ \vdots \\ c_{k,k-1}p\theta_2 \\ c_{k,k}p\theta_2 \end{pmatrix}$$

$$= \frac{1}{1 - q\theta_1\,\theta_2 \sum_{i=0}^{k-1} (p\theta_2)^i} \begin{pmatrix} (p\theta_2)^{k-1} \\ (p\theta_2)^{k-1} \\ c_{k,3}p\theta_2 \\ \vdots \\ c_{k,k}p\theta_2 \end{pmatrix}.$$

On the other hand, since the determinant of the matrix $\mathbf{I} - \mathbf{T}_{k+1}\boldsymbol{\Delta}_{k+1}(\boldsymbol{\theta}^{\mathbf{r}})$ is equal to the determinant of the matrix $\mathbf{A}_k - \mathbf{bc}$, we get

$$\det\left(\mathbf{I} - \mathbf{T}_{k+1}\boldsymbol{\Delta}_{k+1}(\boldsymbol{\theta}^{\mathbf{r}})\right) = 1 - q\theta_1\,\theta_2 \sum_{i=0}^{k-1} (p\theta_2)^i .$$

Consequently, the cofactors $c_{k+1,1}$ and $c_{k+1,2}$ of the matrix $\mathbf{I} - \mathbf{T}_{k+1}\boldsymbol{\Delta}_{k+1}(\boldsymbol{\theta}^{\mathbf{r}})$ are equal to $(p\theta_2)^{k-1}$.

Finally, we conclude that

$$\boldsymbol{\pi}_{k+1}\boldsymbol{\Delta}_{k+1}(\boldsymbol{\theta}^{\mathbf{r}})\Big(\mathbf{I} - \mathbf{T}_{k+1}\boldsymbol{\Delta}_{k+1}(\boldsymbol{\theta}^{\mathbf{r}})\Big)^{-1}\mathbf{t}_{k+1} = \frac{(p\theta_2)^k}{1 - q\theta_1\,\theta_2 \sum\limits_{i=0}^{k-1} (p\theta_2)^i}$$

and $(\boldsymbol{\pi}_{k+1}, \mathbf{T}_{k+1}; \mathbf{R}_{k+1})$ is a MDPH*-representation for the bivariate Geometric distribution of order $k$.

**Important remark.**

We can reduce the dimensionality of the representation in (5.19) by removing the zero-column in the sub-transition probability matrix $\mathbf{T}_{k+1}$, and instead of that we consider the reward of the vector of exits. That is formulated as follows.

Consider the initial vector

$$\mathbf{u}_k = (q, p, 0, \ldots, 0),$$

(an $k$-dimensional row vector), the sub-transition probability matrix

$$\mathbf{V}_k = \begin{pmatrix} q & p & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ q & 0 & 0 & p & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ q & 0 & 0 & 0 & \cdots & p \\ q & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

(an $k$-square matrix), the vector of exits

$$\mathbf{v}_{_k} = (0, 0, 0, \dots, 0, p)^\top ,$$

(an $k$-dimensional column vector), and the matrix of rewards

$$\mathbf{R}_{_k} = \begin{array}{c} \begin{matrix} M_{_0} & T \end{matrix} \\ \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \end{array}.$$

Then, we can prove by induction that

$$\mathbf{u}_k \boldsymbol{\Delta}_k(\boldsymbol{\theta}^{\mathbf{r}}) \Big(\mathbf{I} - \mathbf{V}_k \boldsymbol{\Delta}_k(\boldsymbol{\theta}^{\mathbf{r}})\Big)^{-1} \boldsymbol{\Delta}_k\left(\boldsymbol{\theta}^{\mathbf{r}}\right) \mathbf{v}_k = \frac{(p\theta_2)^k}{1 - q\theta_1\,\theta_2 \sum\limits_{i=0}^{k-1} (p\theta_2)^i} .$$

That means that if we consider the rewards produced by the exit vector, then $(\mathbf{u}_{_k}, \mathbf{V}_{_k}; \mathbf{R}_{_k})$ is another representation for $(M_{_0}, T)$ . Therefore, in general we can find other representations for a multivariate discrete phase-type distribution that are more convenient in practice.

### 5.3.2 Bivariate Negative Binomial distribution of order $k$

**Definition 5.13** The joint distribution of two non-negative random variables $N_1$ and $N_2$ is said to be bivariate negative binomial distributed of order $k$ if their joint probability-generating function is given by

$$\mathbb{E}\left(\theta_1^{N_1}\theta_2^{N_2}\right) = \left(\frac{1 - p_1 - p_2}{1 - \theta_1 p_1 - \theta_2 p_2}\right)^k , \tag{5.23}$$

where $k, p_1, p_2 > 0$ and $p_1 + p_2 < 1$, see [Dun67].

From here it follows that $N_1 \sim \mathrm{NB}(k, \frac{p_1}{1-p_2})$ and $N_2 \sim \mathrm{NB}(k, \frac{p_2}{1-p_1})$.

For the interpretation of this bivariate distribution, let us take a fixed $k \in \mathbb{N}$ and consider a manufacturing line of articles with three different types of items: type 1, 2 and $F$. The number of items of type 1 is described by a random variable $N_1$ (which could also be interpreted as the number of successes of the random variable $N_1$) as well as the number of items of type 2 with the random variable $N_2$, and items of type $F$ refer to the failures. The line process is stopped until $k$ failures appear.

We consider a Markov chain which describes independent trials with three possibilities: the success of $N_1$ with probability $p_1$, the success of $N_2$ with probability $p_2$, and the failure with probability $1 - p_1 - p_2$.

First, let us take the case $k = 2$. This means that the Markov chain stops when the second failure appears.

The initial distribution is given by

$$\boldsymbol{\pi}_2 = \left(p_1, p_2, (1 - p_1 - p_2)\, p_1, (1 - p_1 - p_2)\, p_2\right),$$

with atom at zero of size $1 - \boldsymbol{\pi}\mathbf{e} = (1 - p_1 - p_2)^2$ (which is the probability of the absorbing state).

The sub-transition probability matrix is given by

$$\mathbf{T}_2 = \begin{pmatrix} p_1 & p_2 & (1-p_1-p_2)p_1 & (1-p_1-p_2)p_2 \\ p_1 & p_2 & (1-p_1-p_2)p_1 & (1-p_1-p_2)p_2 \\ 0 & 0 & p_1 & p_2 \\ 0 & 0 & p_1 & p_2 \end{pmatrix}.$$

Thus, the exit vector is

$$\mathbf{t}_2 = \begin{pmatrix} (1-p_1-p_2)^2 \\ (1-p_1-p_2)^2 \\ 1-p_1-p_2 \\ 1-p_1-p_2 \end{pmatrix}.$$

Lastly, the matrix of rewards is given by

$$\mathbf{R}_2 = \begin{array}{c} \begin{array}{cc} N_1 & N_2 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}.$$

Then

$$\begin{aligned} & \mathbb{P}(\tau = 0) + \boldsymbol{\pi}_2 \boldsymbol{\Delta}_2 (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_2 \boldsymbol{\Delta}_2 (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_2 \\ = \;& (1-p_1-p_2)^2 + \boldsymbol{\pi}_2 \boldsymbol{\Delta}_2 (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_2 \boldsymbol{\Delta}_2 (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_2 \end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{1-p_1-p_2}{1-\theta_1 p_1-\theta_2 p_2}\right)^2 \\
&= \mathbb{E}\left(\theta_1^{N_1}\theta_2^{N_2}\right).
\end{aligned}
$$

Therefore, the representation $(\boldsymbol{\pi}_{_2}, \mathbf{T}_{_2}; \mathbf{R}_{_2})$ is a bivariate discrete phase-type representation for the bivariate Negative binomial distribution of order 2.

In order to prove the general case, let us fix $k \in \mathbb{N}$ and consider $(\boldsymbol{\pi}_{_k}, \mathbf{T}_{_k}; \mathbf{R}_{_k})$ , where

$$
\boldsymbol{\pi}_{_k} = (a_{_1}, a_{_2}, \ldots, a_{_{2k}}),
$$

every entry $a_{_i}$, $i = 1, \ldots, 2k$, is given by

$$
a_{_i} = \begin{cases}
(1-p_1-p_2)^{k_i} p_1 & \text{if } i \text{ is odd, and } k_{_i} = i-j, \\
& \quad \text{where } j \text{ is such that } i = 2j-1, \text{ and } j = 1, 2, \ldots, k. \\
(1-p_1-p_2)^{k_i} p_2 & \text{if } i \text{ is even, and } k_{_i} = i-(j+1), \\
& \quad \text{where } j \text{ is such that } i = 2j, \text{ and } j = 1, 2, \ldots, k.
\end{cases}
$$

$$
\mathbf{T}_{_k} = \begin{pmatrix}
\mathbf{T}_{_{k-1}} & \mathbf{v}_{_{2k-1}} & \mathbf{v}_{_{2k}} \\
\mathbf{0} & p_1 & p_2 \\
\mathbf{0} & p_1 & p_2
\end{pmatrix},
$$

where $\mathbf{v}_{_{2k-1}}$ is defined as follows

$$
\mathbf{v}_{_{2k-1}} = \left(t_{_{1,2k-1}}, t_{_{2,2k-1}}, \ldots, t_{_{2k-2,2k-1}}\right)^{\top},
$$

where

$$
t_{i,2k-1} = \begin{cases}
(1-p_1-p_2)^{k_i} p_1 & \text{if } i \text{ is odd, and } k_{_i} = k-j, \\
& \quad \text{where } j \text{ is such that } i = 2j-1, \text{ and } j = 1, 2, \ldots, k, \\
(1-p_1-p_2)^{k_i} p_1 & \text{if } i \text{ is even, and } k_{_i} = k-j, \\
& \quad \text{where } j \text{ is such that } i = 2j, \text{ and } j = 1, 2, \ldots, k,
\end{cases}
$$

for every $i = 1, \ldots, 2k$.

The vector $\mathbf{v}_{_{2k}}$ is defined as

$$
\mathbf{v}_{_{2k}} = \left(t_{_{1,2k}}, t_{_{2,2k}}, \ldots, t_{_{2k-2,2k}}\right)^{\top},
$$

where for every $i = 1, \ldots, 2k$, we have

$$
t_{i,2k} = \begin{cases}
(1-p_1-p_2)^{k_i} p_2 & \text{if } i \text{ is odd, and } k_{_i} = k-j, \\
& \quad \text{where } j \text{ is such that } i = 2j-1, \text{ and } j = 1, 2, \ldots, k. \\
(1-p_1-p_2)^{k_i} p_2 & \text{if } i \text{ is even, and } k_{_i} = k-j, \\
& \quad \text{where } j \text{ is such that } i = 2j, \text{ and } j = 1, 2, \ldots, k.
\end{cases}
$$

Notice that

$$\mathbf{v}_{2k-1} = \mathbf{t}_{k-1} p_1 \quad \text{and} \quad \mathbf{v}_{2k} = \mathbf{t}_{k-1} p_2.$$

The exit vector is given by

$$\mathbf{t}_k = \left(\mathbf{t}_{k-1}(1-p_1-p_2), 1-p_1-p_2, 1-p_1-p_2\right)^\top .$$

Lastly, the matrix of rewards

$$\mathbf{R}_k = \begin{array}{c} \begin{matrix} N_1 & N_2 \end{matrix} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array},$$

is an $2k \times 2$-dimensional matrix.

In the next, we are going to calculate

$$\boldsymbol{\pi}_k \boldsymbol{\Delta}_k(\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_k \boldsymbol{\Delta}_k(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_k .$$

**Induction hypothesis.** Assume that

$$\boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1}(\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1}(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_{k-1} = \left(\frac{1-p_1-p_2}{1-\theta_1 p_1 - \theta_2 p_2}\right)^{k-1} - (1-p_1-p_2)^{k-1}. \tag{5.24}$$

Now, we calculate

$$\left(\mathbf{I} - \mathbf{T}_k \boldsymbol{\Delta}_k(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} =$$

$$\begin{pmatrix} \left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1}(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} & a\left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1}(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_{k-1} & b\left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1}(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_{k-1} \\ \mathbf{0} & \frac{1-\theta_2 p_2}{1-\theta_1 p_1 - \theta_2 p_2} & \frac{\theta_2 p_2}{1-\theta_1 p_1 - \theta_2 p_2} \\ \mathbf{0} & \frac{\theta_1 p_1}{1-\theta_1 p_1 - \theta_2 p_2} & \frac{1-\theta_1 p_1}{1-\theta_1 p_1 - \theta_2 p_2} \end{pmatrix},$$

where the constants

$$a = \frac{(1-\theta_2 p_2)\theta_1 p_1 + \theta_1 \theta_2 p_1 p_2}{1-\theta_1 p_1 - \theta_2 p_2}, \qquad b = \frac{(1-\theta_1 p_1)\theta_2 p_2 + \theta_1 \theta_2 p_1 p_2}{1-\theta_1 p_1 - \theta_2 p_2}.$$

We also observe that

$$\begin{aligned}
\boldsymbol{\pi}_k \boldsymbol{\Delta}_k (\boldsymbol{\theta}^{\mathbf{r}}) &= \left(\boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}}), a_{2k-1}\theta_1, a_{2k}\theta_2\right) \\
&= \left(\boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}}), (1-p_1-p_2)^{k-1} p_1\theta_1, (1-p_1-p_2)^{k-1} p_2\theta_2\right).
\end{aligned}$$

Then, by calculating

$$\boldsymbol{\pi}_k \boldsymbol{\Delta}_k (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_k \boldsymbol{\Delta}_k (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} =$$

$$\begin{pmatrix}
\boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \\
a\boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_{k-1} + \frac{p_1\theta_1(1-\theta_2p_2)(1-p_1-p_2)^{k-1}}{1-\theta_1p_1-\theta_2p_2} + \frac{p_1p_2\theta_2\theta_1(1-p_1-p_2)^{k-1}}{1-\theta_1p_1-\theta_2p_2} \\
b\boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_{k-1} + \frac{p_1p_2\theta_1\theta_2(1-p_1-p_2)^{k-1}}{1-\theta_1p_1-\theta_2p_2} + \frac{(1-\theta_1p_1)p_2\theta_2(1-p_1-p_2)^{k-1}}{1-\theta_1p_1-\theta_2p_2}
\end{pmatrix}^{\top}$$

and consequently

$$\boldsymbol{\pi}_k \boldsymbol{\Delta}_k (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_k \boldsymbol{\Delta}_k (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_k =$$

$$\begin{aligned}
& (1-p_1-p_2)\boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_{k-1} \\
+ \quad & (1-p_1-p_2)\left(\frac{(1-\theta_2p_2)\theta_1p_1 + \theta_1\theta_2p_1p_2}{1-\theta_1p_1-\theta_2p_2}\right) \boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_{k-1} \\
+ \quad & \frac{p_1\theta_1(1-\theta_2p_2)(1-p_1-p_2)^k}{1-\theta_1p_1-\theta_2p_2} + \frac{p_1p_2\theta_2\theta_1(1-p_1-p_2)^k}{1-\theta_1p_1-\theta_2p_2} \\
+ \quad & (1-p_1-p_2)\left(\frac{(1-\theta_1p_1)\theta_2p_2 + \theta_1\theta_2p_1p_2}{1-\theta_1p_1-\theta_2p_2}\right) \boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_{k-1} \\
+ \quad & \frac{p_1p_2\theta_1\theta_2(1-p_1-p_2)^k}{1-\theta_1p_1-\theta_2p_2} + \frac{(1-\theta_1p_1)p_2\theta_2(1-p_1-p_2)^k}{1-\theta_1p_1-\theta_2p_2} \\
= \quad & \left(\frac{1-p_1-p_2}{1-\theta_1p_1-\theta_2p_2}\right) \boldsymbol{\pi}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}}) \left(\mathbf{I} - \mathbf{T}_{k-1} \boldsymbol{\Delta}_{k-1} (\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1} \mathbf{t}_{k-1} + \frac{(\theta_1p_1+\theta_2p_2)(1-p_1-p_2)^k}{1-\theta_1p_1-\theta_2p_2} \\
= \quad & \left(\frac{1-p_1-p_2}{1-\theta_1p_1-\theta_2p_2}\right)\left\{\left(\frac{1-p_1-p_2}{1-\theta_1p_1-\theta_2p_2}\right)^{k-1} - (1-p_1-p_2)^{k-1}\right\} + \frac{(\theta_1p_1+\theta_2p_2)(1-p_1-p_2)^k}{1-\theta_1p_1-\theta_2p_2} \\
= \quad & \left(\frac{1-p_1-p_2}{1-\theta_1p_1-\theta_2p_2}\right)^k - \left\{\frac{1-\theta_1p_1-\theta_2p_2}{1-\theta_1p_1-\theta_2p_2} + \right\}(1-p_1-p_2)^k \\
= \quad & \left(\frac{1-p_1-p_2}{1-\theta_1p_1-\theta_2p_2}\right)^k - (1-p_1-p_2)^k.
\end{aligned}$$

Finally, we can conclude that

$$(1-p_1-p_2)^k + \boldsymbol{\pi}_{k}\boldsymbol{\Delta}_{k}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\left(\mathbf{I}-\mathbf{T}_{k}\boldsymbol{\Delta}_{k}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\right)^{-1}\mathbf{t}_{k} = \left(\frac{1-p_1-p_2}{1-\theta_1 p_1-\theta_2 p_2}\right)^k.$$

**Remark.**

We can generalise the idea of the bivariate discrete phase-type representation for the bivariate Negative binomial to construct MDPH*-representations for Negative multinomial distributions. The probability-generating function of a negative multinomial distribution with parameters $(k; p_0, p_1, \ldots, p_r)$ is

$$g(\theta_1, \ldots, \theta_r) = p_0^k \left(1 - \sum_{i=1}^{r} \theta_i p_i\right)^{-k},$$

where $p_0 = 1 - \sum_{i=1}^{r} p_i$.

### 5.3.3 Edwards-Gurland bivariate negative binomial (Compound correlated bivariate Poisson distribution)

Let $(W_1, W_2)$ be a pair of correlated random variables. Then $(W_1, W_2)$ is said to be the Edwards-Gurland bivariate negative binomial if its joint probability-generating function is

$$\mathbb{E}\left(\theta_1^{W_1}\theta_2^{W_2}\right) = \left(\frac{1-p_1-p_2-p_3}{1-p_1\theta_1-p_2\theta_2-p_3\theta_1\theta_2}\right)^k,$$

where $k \in \mathbb{N}$, $p_i > 0, i = 1, 2, 3$, and $p_1 + p_2 + p_3 < 1$, (see [EG61]).

Note that if $p_3 = 0$, then the Edwards-Gurland bivariate negative binomial is a bivariate negative binomial (see Equation (5.23)).

For $k = 2$, a MDPH*-representation is given by $(\boldsymbol{\pi}_2, \mathbf{T}_2; \mathbf{R}_2)$, where

$$\boldsymbol{\pi}_2 = (p_1, p_2, p_3, (1-p_1-p_2-p_3)p_1, (1-p_1-p_2-p_3)p_2, (1-p_1-p_2-p_3)p_3),$$

which is the initial probability distribution of the underlying Markov chain with an atom at zero of size $(1-p_1-p_2-p_3)^2$.

The sub-transition probability matrix is

$$\mathbf{T}_2 = \begin{pmatrix} p_1 & p_2 & p_3 & (1-p_1-p_2-p_3)p_1 & (1-p_1-p_2-p_3)p_2 & (1-p_1-p_2-p_3)p_3 \\ p_1 & p_2 & p_3 & (1-p_1-p_2-p_3)p_1 & (1-p_1-p_2-p_3)p_2 & (1-p_1-p_2-p_3)p_3 \\ p_1 & p_2 & p_3 & (1-p_1-p_2-p_3)p_1 & (1-p_1-p_2-p_3)p_2 & (1-p_1-p_2-p_3)p_3 \\ 0 & 0 & 0 & p_1 & p_2 & p_3 \\ 0 & 0 & 0 & p_1 & p_2 & p_3 \\ 0 & 0 & 0 & p_1 & p_2 & p_3 \end{pmatrix},$$

and the exit vector is

$$\mathbf{t}_2 = \begin{pmatrix} (1-p_1-p_2-p_3)^2 \\ (1-p_1-p_2-p_3)^2 \\ (1-p_1-p_2-p_3)^2 \\ 1-p_1-p_2-p_3 \\ 1-p_1-p_2-p_3 \\ 1-p_1-p_2-p_3 \end{pmatrix}.$$

The matrix of rewards is

$$\begin{array}{cc} & \begin{array}{cc} W_1 & W_2 \end{array} \\ \mathbf{R}_2 = & \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}. \end{array}$$

In general we can form the initial vector recursively as is shown next.

$$\boldsymbol{\pi}_k = \left(\boldsymbol{\pi}_{k-1}, (1-p_1-p_2-p_3)^{k-1}p_1, (1-p_1-p_2-p_3)^{k-1}p_2, (1-p_1-p_2-p_3)^{k-1}p_3\right), \tag{5.25}$$

which corresponds to a vector of dimension $3k$.

The sub-transition probability matrix is going to be given by

$$\mathbf{T}_k = \begin{pmatrix} \mathbf{T}_{k-1} & p_1\mathbf{t}_{k-1} & p_2\mathbf{t}_{k-1} & p_3\mathbf{t}_{k-1} \\ \mathbf{0} & p_1 & p_2 & p_3 \\ \mathbf{0} & p_1 & p_2 & p_3 \\ \mathbf{0} & p_1 & p_2 & p_3 \end{pmatrix}, \tag{5.26}$$

which is a $3k$-square matrix.

The exit vector is

$$\mathbf{t}_k = \begin{pmatrix} \mathbf{t}_{k-1}(1-p_1-p_2-p_3) \\ 1-p_1-p_2-p_3 \\ 1-p_1-p_2-p_3 \\ 1-p_1-p_2-p_3 \end{pmatrix}$$

which is of dimension $3k$.

Lastly, the matrix of rewards is given by

$$\mathbf{R}_{_k} = \mathbf{e}^{\top} \otimes \mathbf{C}, \tag{5.27}$$

where $\mathbf{e}^{\top}$ is the column vector of ones of dimension $k$ and

$$\mathbf{C} = \begin{array}{c} \begin{matrix} W_1 & W_2 \end{matrix} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \end{array}. \tag{5.28}$$

The matrix of rewards $\mathbf{R}_{_k}$ is of dimension $3k \times 2$.

We can prove that

$$(\boldsymbol{\pi}_{_k}, \mathbf{T}_{_k}; \mathbf{R}_{_k}),$$

which is defined in Equations (5.25), (5.26) and (5.27), respectively, is a MDPH*-representation for the distribution of $(W_1, W_2)$ with a zero point of size $(1 - p_1 - p_2 - p_3)^k$. The proof is similar than in Example 5.3.2.

### 5.3.4 Type I: Method of trivariate reduction or random element in common: bivariate case.

Let $Z, W_1$ and $W_2$ be independent and geometric distributed random variables with parameter $p, p_1$ and $p_2$, respectively. Let $Y_i = Z + W_i$, $i = 1, 2$.

The probability-generating function is obtained as follows.

$$\begin{aligned}
\mathbb{E}\left(\theta_1^{Y_1}\theta_2^{Y_2}\right) &= \mathbb{E}\left(\theta_1^{Z+W_1}\theta_2^{Z+W_2}\right) \\
&= \mathbb{E}\left((\theta_1\theta_2)^Z\theta_1^{W_1}\theta_2^{W_2}\right) \\
&= \mathbb{E}\left((\theta_1\theta_2)^Z\right)\mathbb{E}\left(\theta_1^{W_1}\right)\mathbb{E}\left(\theta_2^{W_2}\right) \\
&= \left(\frac{p}{1-\theta_1\theta_2 q}\right)\left(\frac{p_1}{1-\theta_1 q_1}\right)\left(\frac{p_2}{1-\theta_2 q_2}\right).
\end{aligned}$$

Consider a Markov chain with initial distribution given by

$$\boldsymbol{\pi} = (q, pq_1, pp_1q_2)$$

and the initial probability of the absorbing state to be $pp_1p_2$.

Define a sub-transition probability matrix given by

$$\mathbf{T} = \begin{pmatrix} q & pq_1 & pp_1q_2 \\ 0 & q_1 & p_1q_2 \\ 0 & 0 & q_2 \end{pmatrix}.$$

Thus, the exit vector is

$$\mathbf{t} = \begin{pmatrix} pp_1p_2 \\ p_1p_2 \\ p_2 \end{pmatrix}.$$

Consider the matrix of rewards given by

$$\mathbf{R} = \begin{array}{c} \begin{matrix} Y_1 & Y_2 \end{matrix} \\ \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}.$$

Then, we conclude that $(\boldsymbol{\pi}, \mathbf{T}; \mathbf{R})$ is a MDPH*-representation for $(Y_1, Y_2)$ by making the following calculations

$$\begin{aligned} \mathbb{P}(\tau = 0) + \boldsymbol{\pi}\boldsymbol{\Delta}(\boldsymbol{\theta}^{\mathbf{r}})\left(\mathbf{I} - \mathbf{T}\boldsymbol{\Delta}(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1}\mathbf{t} &= pp_1p_2 + \boldsymbol{\pi}\boldsymbol{\Delta}(\boldsymbol{\theta}^{\mathbf{r}})\left(\mathbf{I} - \mathbf{T}\boldsymbol{\Delta}(\boldsymbol{\theta}^{\mathbf{r}})\right)^{-1}\mathbf{t} \\ &= \left(\frac{p}{1-\theta_1\theta_2 q}\right)\left(\frac{p_1}{1-\theta_1 q_1}\right)\left(\frac{p_2}{1-\theta_2 q_2}\right) \\ &= \mathbb{E}\left(\theta_1^{Y_1}\theta_2^{Y_2}\right). \end{aligned}$$

### 5.3.5 Extended trivariate reduction: bivariate case.

The method of trivariate reduction given in Example 5.3.4 can be extended by defining the variables

$$X_1 = Y_1 + W_1, \quad X_2 = Y_2 + W_2,$$

where $Y_1$ and $Y_2$ are independent and Negative binomial distributed with parameters $(k_1, \frac{p_1+p_3}{1-p_2})$ and $(k_2, \frac{p_2+p_3}{1-p_1})$, respectively, while $(W_1, W_2)$ is a pair of Edwards-Gurland bivariate Negative binomial with parameters $(k_3, p_1, p_2, p_3)$, and it is independent of the variables $Y_1$ and $Y_2$ (see [MNOS10]).

The probability-generating function for $(X_1, X_2)$ is derived next.

$$\mathbb{E}\left(\theta_1^{X_1}\theta_2^{X_2}\right) = \mathbb{E}\left(\theta_1^{Y_1+W_1}\theta_2^{Y_2+W_2}\right)$$

$$
\begin{aligned}
&= \mathbb{E}\left(\theta_1^{Y_1}\right)\mathbb{E}\left(\theta_2^{Y_2}\right)\mathbb{E}\left(\theta_1^{W_1}\theta_2^{W_2}\right) \\
&= \left(\frac{d_1}{1-b_1\theta_1}\right)^{k_1}\left(\frac{d_2}{1-b_2\theta_2}\right)^{k_2}\left(\frac{1-p_1-p_2-p_3}{1-p_1\theta_1-p_2\theta_2-p_3\theta_1\theta_2}\right)^{k_3},
\end{aligned}
$$

where $b_1 = \frac{p_1+p_3}{1-p_2}, b_2 = \frac{p_2+p_3}{1-p_1}$ and $d_i = 1-b_i, i = 1, 2$.

In order to obtain one MDPH*-representation for $(X_1, X_2)$, we are going to use the MDPH*-representations for $(W_1, W_2)$ given in Equation (5.3.3).

Consider a Markov chain with initial distribution given by

$$
\boldsymbol{\pi} = \left(\boldsymbol{\pi}_{k_3}, \epsilon_{k_3} b_1, \epsilon_{k_3} d_1 b_1, \ldots, \epsilon_{k_3} d_1^{k_1-1} b_1, \epsilon_{k_3} d_1^{k_1} b_2, \ldots, \epsilon_{k_3} d_1^{k_1} d_2^{k_2-1} b_2\right),
$$

which is a row vector of dimension $3k_3 + k_1 + k_2$, where $\boldsymbol{\pi}_{k_3}$ is described in Equation (5.25) and $\epsilon_{k_3} = (1-p_1-p_2-p_3)^{k_3}$. The initial probability of the absorbing state is $a_k = \epsilon_{k_3} d_1^{k_1} d_2^{k_2}$.

The sub-transition probability matrix $\mathbf{T}$ of the Markov chain is

$$
\begin{pmatrix}
\mathbf{T}_{k_3} & \mathbf{t}_{k_3} b_1 & \mathbf{t}_{k_3} d_1 b_1 & \mathbf{t}_{k_3} d_1^2 b_1 & \cdots & \mathbf{t}_{k_3} d_1^{k_1-1} b_1 & \mathbf{t}_{k_3} d_1^{k_1} b_2 & \mathbf{t}_{k_3} d_1^{k_1} d_2 b_2 & \cdots & \mathbf{t}_{k_3} d_1^{k_1} d_2^{k_2-1} b_2 \\
\mathbf{0} & b_1 & d_1 b_1 & d_1^2 b_1 & \cdots & d_1^{k_1-1} b_1 & d_1^{k_1} b_2 & d_1^{k_1} d_2 b_2 & \cdots & d_1^{k_1} d_2^{k_2-1} b_2 \\
\mathbf{0} & 0 & b_1 & d_1 b_1 & \cdots & d_1^{k_1-2} b_1 & d_1^{k_1-1} b_2 & d_1^{k_1-1} d_2 b_2 & \cdots & d_1^{k_1-1} d_2^{k_2-2} b_2 \\
\mathbf{0} & 0 & 0 & b_1 & \cdots & d_1^{k_1-3} b_1 & d_1^{k_1-2} b_2 & d_1^{k_1-2} d_2 b_2 & \cdots & d_1^{k_1-2} d_2^{k_2-3} b_2 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & 0 & 0 & 0 & \cdots & b_1 & d_1 b_2 & d_1 d_2 b_2 & \cdots & d_1 d_2^{k_2-1} b_2 \\
\mathbf{0} & 0 & 0 & 0 & \cdots & 0 & b_2 & d_2 b_2 & \cdots & d_2^{k_2-2} b_2 \\
\mathbf{0} & 0 & 0 & 0 & \cdots & 0 & 0 & b_2 & \cdots & d_2^{k_2-3} b_2 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & b_2
\end{pmatrix},
$$

which is a square matrix of dimension $3k_3 + k_1 + k_2$.

Lastly, we consider the matrix of rewards given by

$$
\mathbf{R} = \begin{pmatrix} \mathbf{e}_{k_3}^{\top} \otimes \mathbf{C} \\ \mathbf{e}_{k_1}^{\top} \otimes \mathbf{v} \\ \mathbf{e}_{k_2}^{\top} \otimes \mathbf{u} \end{pmatrix},
$$

where the matrix $\mathbf{C}$ is described in Equation (5.28), while $\mathbf{v} = (1, 0)$ and $\mathbf{u} = (0, 1)$.

### 5.3.6 Type II: Decomposition.

Let $J_i \sim$ Bernoulli $(1 - \beta_i)$, $W_i \sim$ Geometric $(D_i)$ and $Z_i \sim$ Geometric $(1 - p_i)$, $i = 1, 2$, where all of these variables are independent. Now, consider the variables

$$Y_i = W_i + J_i Z_i, \quad i = 1, 2.$$

#### 5.3.6.1 Case 1: $W_1, W_2 \sim$ Geometric $(D)$.

The probability-generating function is

$$\begin{aligned} & \mathbb{E}\left(\theta_1^{Y_1}\theta_2^{Y_2}\right) \\ = & \ \mathbb{E}\left((\theta_1\theta_2)^W\right)\mathbb{E}\left(\theta_1^{J_1 Z_1}\right)\mathbb{E}\left(\theta_2^{J_2 Z_2}\right) \\ = & \left(\frac{D}{1-\theta_1\theta_2 d}\right)\left(\frac{(1-\beta_1)p_1 + \beta_1(1-\theta_1 q_1)}{1-\theta_1 q_1}\right)\left(\frac{(1-\beta_2)p_2 + \beta_2(1-\theta_2 q_2)}{1-\theta_2 q_2}\right), \end{aligned}$$

where $d = 1 - D$ and $q_i = 1 - p_i, i = 1, 2$.

Consider a Markov chain with initial distribution

$$\boldsymbol{\pi} = (d, D\beta_1 q_1, D(1-\beta_1)\beta_2 q_2 + D\beta_1 p_1 \beta_2 q_2)$$

and with initial probability of the absorbing state

$$D\beta_1 p_1(1-\beta_2) + D(1-\beta_1)(1-\beta_2) + D(1-\beta_1)\beta_2 p_2 + D\beta_1 p_1 \beta_2 p_2.$$

The sub-transition probability matrix is given by

$$\mathbf{T} = \begin{pmatrix} d & D\beta_1 q_1 & D(1-\beta_1)\beta_2 q_2 + D\beta_1 p_1 \beta_2 q_2 \\ 0 & q_1 & p_1 \beta_2 q_2 \\ 0 & 0 & q_2 \end{pmatrix}.$$

Then, the exit vector is

$$\mathbf{t} = \begin{pmatrix} D\beta_1 p_1(1-\beta_2) + D\beta_1 p_1 \beta_2 p_2 + D(1-\beta_1)(1-\beta_2) + D(1-\beta_1)\beta_2 p_2 \\ p_1(1-\beta_2) + p_1\beta_2 p_2 \\ p_2 \end{pmatrix}.$$

Finally, the matrix of rewards is given by

$$\mathbf{R} = \begin{array}{c} \begin{array}{cc} Y_1 & Y_2 \end{array} \\ \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}.$$

#### 5.3.6.2 Case 2: sharing the variable $JZ$.

Let $J \sim$ Bernoulli $(\beta)$ , $Z \sim$ Geometric $(p)$ and $W_i \sim$ Geometric $(p_i)$ for $i = 1, 2$, and all of these variables are independent.

The probability-generating function is given by

$$\begin{aligned} & \mathbb{E}\left(\theta_1^{Y_1}\theta_2^{Y_2}\right) \\ = \; & \mathbb{E}\left(\theta_1^{W_1}\right)\mathbb{E}\left(\theta_2^{W_2}\right)\mathbb{E}\left((\theta_1\theta_2)^{JZ}\right) \\ = \; & \left(\frac{p_1}{1-(1-p_1)\theta_1}\right)\left(\frac{p_2}{1-(1-p_2)\theta_2}\right)\left(\frac{\beta p + (1-\beta)\left(1-(1-p)\theta_1\theta_2\right)}{1-(1-p)\theta_1\theta_2}\right). \end{aligned}$$

Consider a Markov chain with initial distribution

$$\boldsymbol{\pi} = (\beta q \, , \beta p q_1 + (1-\beta) q_1 , \beta p p_1 q_2 + (1-\beta) p_1 q_2) \, ,$$

where $q_i = 1 - p_i, i = 1, 2$, and the initial probability of the absorbing state is

$$\beta p p_1 p_2 + (1-\beta) p_1 p_2.$$

The sub-transition probability matrix is given by

$$\mathbf{T} = \begin{pmatrix} q & pq_1 & pp_1q_2 \\ 0 & q_1 & p_1q_2 \\ 0 & 0 & q_2 \end{pmatrix}.$$

Thus, the exit vector is

$$\mathbf{t} = \begin{pmatrix} pp_1p_2 \\ p_1p_2 \\ p_2 \end{pmatrix}.$$

Lastly, the matrix of rewards is given by

$$\mathbf{R} = \begin{array}{c} \begin{array}{cc} Y_1 & Y_2 \end{array} \\ \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}.$$

### 5.3.7 Type III: N-fold of geometric distributions.

Let $N \sim$ Geometric$(1-d)$ taking values in $\{0,1,2,\ldots\}$ and consider the variables given by

$$Y_i = \sum_{j=1}^{N} Z_{i,j}, \quad i = 1,2,$$

where $Z_{i,j} \sim$ Geometric$(p_i), i = 1,2.$

The joint probability-generating function of $(Y_1, Y_2)$ is derived as follows

$$\begin{aligned}
\mathbb{E}\left(\theta_1^{Y_1}\theta_2^{Y_2}\right) &= \sum_{n=0}^{\infty} \mathbb{P}\left(N=n\right)\mathbb{E}\left(\theta_1^{Y_1}|N=n\right)\mathbb{E}\left(\theta_2^{Y_2}|N=n\right) \\
&= d\sum_{n=0}^{\infty}(1-d)^n\left(\frac{p_1}{1-q_1\theta_1}\right)^n\left(\frac{p_2}{1-q_2\theta_2}\right)^n \\
&= d\sum_{n=0}^{\infty}\left(\frac{(1-d)p_1p_2}{(1-q_1\theta_1)(1-q_2\theta_2)}\right)^n \\
&= \frac{d(1-q_1\theta_1)(1-q_2\theta_2)}{(1-q_1\theta_1)(1-q_2\theta_2)-(1-d)p_1p_2}.
\end{aligned}$$

The first equality holds due to $Y_1$ and $Y_2$ are conditionally independent given $N$.

Let us consider a Markov chain with initial distribution

$$\boldsymbol{\pi} = \left(dq_1 + \frac{dp_1p_2dq_1}{1-dp_1p_2}, dp_1q_2 + \frac{dp_1p_2dp_1q_2}{1-dp_1p_2}\right),$$

where $q_i = 1 - p_i, i = 1,2,$ and the initial probability of the absorbing state is

$$(1-d) + \frac{dp_1p_2(1-d)}{1-dp_1p_2}.$$

The sub-transition probability matrix is given by

$$\mathbf{T} = \begin{pmatrix} q_1 + \dfrac{p_1p_2dq_1}{1-dp_1p_2} & p_1q_2 + \dfrac{p_1p_2dp_1q_2}{1-dp_1p_2} \\ \\ \dfrac{p_2dq_1}{1-dp_1p_2} & q_2 + \dfrac{p_2dp_1q_2}{1-dp_1p_2} \end{pmatrix}.$$

Then, the exit vector is

$$\mathbf{t} = \begin{pmatrix} \dfrac{p_1p_2(1-d)}{1-dp_1p_2} \\ \\ \dfrac{p_2(1-d)}{1-dp_1p_2} \end{pmatrix}.$$

Lastly, the matrix of rewards is

$$\mathbf{R} = \begin{array}{c} \begin{matrix} Y_1 & Y_2 \end{matrix} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}.$$

### 5.3.8 Joint order statistics from DPH-distributions

Here, we only consider the joint distribution of the minimum and maximum from DPH-distributions. We are going to calculate the joint probability-generating function of the minimum and the maximum of two independent and DPH-distributed random variables.

Let $Y_1$ and $Y_2$ be independent and DPH-distributed with representation given by $(\boldsymbol{\pi}_1, \mathbf{T}_1)$ and $(\boldsymbol{\pi}_2, \mathbf{T}_2)$ , respectively.

Denote $Y_{(1:2)} = \min(Y_1, Y_2)$ and $Y_{(2:2)} = \max(Y_1, Y_2)$ and recall the DPH-representation for the maximum $Y_{(2:2)}$, this is $\left((\overline{\boldsymbol{\pi}}_2, \mathbf{0}), \mathbf{T}_{(2)}\right)$ (see Equation (4.29)), where

$$\begin{aligned}
(\overline{\boldsymbol{\pi}}_2, \mathbf{0}) &= (\boldsymbol{\pi}_1 \otimes \boldsymbol{\pi}_2, 0, 0), \\
\mathbf{T}_{(2)} &= \begin{pmatrix} \mathbf{T}_1 \otimes \mathbf{T}_2 & \mathbf{T}_1 \otimes \mathbf{t}_2 & \mathbf{t}_1 \otimes \mathbf{T}_2 \\ \mathbf{0} & \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_2 \end{pmatrix},
\end{aligned}$$

where $\mathbf{t}_i = \mathbf{e} - \mathbf{T}_i\mathbf{e}, i = 1, 2.$

The vector of exits of $\mathbf{T}_{(2)}$ is given by

$$\mathbf{t}_{(2)} = (\mathbf{t}_1 \otimes \mathbf{t}_2, \mathbf{t}_1, \mathbf{t}_2)^\top.$$

Let $\theta_1, \theta_2 \in (0, 1)$ . Then

$$\begin{aligned}
&\mathbb{E}\left(\theta_1^{Y_{(1:2)}} \theta_2^{Y_{(2:2)}}\right) \\
=& \sum_{m_1, m_2 \geq 1} \theta_1^{m_1} \theta_2^{m_2} \mathbb{P}\left(Y_{1:2} = m_1, Y_{2:2} = m_2\right) \\
=& \sum_{m_2 \geq 1} \sum_{m_1 = 1}^{m_2} \theta_1^{m_1} \theta_2^{m_2} \mathbb{P}\left(Y_{1:2} = m_1, Y_{2:2} = m_2\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{m_2\geq 1} \theta_2^{m_2}\theta_1 \sum_{m_1=1}^{m_2} \theta_1^{m_1-1}\mathbb{P}\left(Y_{1:2}=m_1, Y_{2:2}=m_2\right)\\
&= \sum_{m_2\geq 1} \theta_2^{m_2}\theta_1 \left(\overline{\boldsymbol{\pi}}_2,\mathbf{0},\mathbf{0}\right)\begin{pmatrix} \mathbf{T}_1\otimes\mathbf{T}_2\theta_1 & \mathbf{T}_1\otimes\mathbf{t}_2 & \mathbf{t}_1\otimes\mathbf{T}_2\\ \mathbf{0} & \mathbf{T}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{T}_2\end{pmatrix}^{m_2-1}\mathbf{t}_{(2)}\\
&= \sum_{m_2\geq 1} \theta_2^{m_2}\theta_1 \left(\overline{\boldsymbol{\pi}}_2,\mathbf{0},\mathbf{0}\right)\\
&\times \begin{pmatrix} \left(\mathbf{A}_{1,2}(\theta_1)\right)^{m_2-1} & \sum_{m=0}^{m_2-2}\left(\mathbf{A}_{1,2}(\theta_1)\right)^{m_2-2-m}\mathbf{B}_{1,2}\mathbf{T}_1^m & \sum_{m=0}^{m_2-2}\left(\mathbf{A}_{1,2}(\theta_1)\right)^{m_2-2-m}\mathbf{B}_{2,1}\mathbf{T}_2^m\\ \mathbf{0} & \mathbf{T}_1^{m_2-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{T}_2^{m_2-1}\end{pmatrix}\mathbf{t}_{(2)},
\end{aligned}
$$

where $\mathbf{A}_{1,2}(\theta_1)=\mathbf{T}_1\otimes\mathbf{T}_2\theta_1, \mathbf{B}_{1,2}=\mathbf{T}_1\otimes\mathbf{t}_2$ and $\mathbf{B}_{2,1}=\mathbf{t}_1\otimes\mathbf{T}_2$.

By continuing with the calculations we get

$$
\begin{aligned}
\mathbb{E}\left(\theta_1^{Y_{(1:2)}}\theta_2^{Y_{(2:2)}}\right) &= \sum_{m_2\geq 1}\theta_2^{m_2}\theta_1\overline{\boldsymbol{\pi}}_2\left(\mathbf{A}_{1,2}(\theta_1)\right)^{m_2-1}\left(\mathbf{t}_1\otimes\mathbf{t}_2\right) && (5.29)\\
&+ \sum_{m_2\geq 1}\theta_2^{m_2}\theta_1\overline{\boldsymbol{\pi}}_2\sum_{m=0}^{m_2-2}\left(\mathbf{A}_{1,2}(\theta_1)\right)^{m_2-2-m}\mathbf{B}_{1,2}\mathbf{T}_1^m\mathbf{t}_1 && (5.30)\\
&+ \sum_{m_2\geq 1}\theta_2^{m_2}\theta_1\overline{\boldsymbol{\pi}}_2\sum_{m=0}^{m_2-2}\left(\mathbf{A}_{1,2}(\theta_1)\right)^{m_2-2-m}\mathbf{B}_{2,1}\mathbf{T}_2^m\mathbf{t}_2. && (5.31)
\end{aligned}
$$

From here, the first sum (5.29) is

$$
\begin{aligned}
&\sum_{m_2\geq 1}\theta_2^{m_2}\theta_1\overline{\boldsymbol{\pi}}_2\left(\mathbf{A}_{1,2}(\theta_1)\right)^{m_2-1}\left(\mathbf{t}_1\otimes\mathbf{t}_2\right)\\
&= \theta_1\theta_2\overline{\boldsymbol{\pi}}_2\sum_{m_2\geq 1}\left(\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{m_2-1}\left(\mathbf{t}_1\otimes\mathbf{t}_2\right)\\
&= \theta_1\theta_2\overline{\boldsymbol{\pi}}_2\left(\mathbf{I}-\left(\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)\right)^{-1}\left(\mathbf{t}_1\otimes\mathbf{t}_2\right). \qquad (5.32)
\end{aligned}
$$

The second sum (5.30) is

$$
\sum_{m_2\geq 1}\theta_2^{m_2}\theta_1\overline{\boldsymbol{\pi}}_2\sum_{m=0}^{m_2-2}\left(\mathbf{A}_{1,2}(\theta_1)\right)^{m_2-2-m}\mathbf{B}_{1,2}\mathbf{T}_1^m\mathbf{t}_1
$$

$$
\begin{aligned}
&= \theta_2^2\theta_1\overline{\boldsymbol{\pi}}_2 \sum_{m_2\geq 2}\sum_{m=0}^{m_2-2} \left(\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{m_2-2-m}\left(\mathbf{T}_1\otimes\mathbf{t}_2\right)\left(\mathbf{T}_1\theta_2\right)^m\mathbf{t}_1\\
&= \theta_2^2\theta_1\overline{\boldsymbol{\pi}}_2 \sum_{u\geq 0}\left(\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{u}\sum_{m\geq 0}\left(\mathbf{T}_1\otimes\mathbf{t}_2\right)\left(\mathbf{T}_1\theta_2\right)^m\mathbf{t}_1\\
&= \theta_2^2\theta_1\overline{\boldsymbol{\pi}}_2\left(\mathbf{I}-\left(\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)\right)^{-1}\left(\mathbf{T}_1\otimes\mathbf{t}_2\right)\left(\mathbf{I}-\left(\mathbf{T}_1\theta_2\right)\right)^{-1}\mathbf{t}_1. \qquad (5.33)
\end{aligned}
$$

The third sum (5.31) is similarly calculated as the second one, thus

$$
\begin{aligned}
&\sum_{m_2\geq 1}\theta_2^{m_2}\theta_1\overline{\boldsymbol{\pi}}_2\sum_{m=0}^{m_2-2}\left(\mathbf{A}_{1,2}(\theta_1)\right)^{m_2-2-m}\mathbf{B}_{2,1}\mathbf{T}_2^m\mathbf{t}_2\\
&= \theta_2^2\theta_1\overline{\boldsymbol{\pi}}_2\left(\mathbf{I}-\left(\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)\right)^{-1}\left(\mathbf{t}_1\otimes\mathbf{T}_2\right)\left(\mathbf{I}-\left(\mathbf{T}_2\theta_2\right)\right)^{-1}\mathbf{t}_2. \qquad (5.34)
\end{aligned}
$$

Finally, from Equations (5.32), (5.33) and (5.34), the joint probability-generating function is given by

$$
\begin{aligned}
\mathbb{E}\left(\theta_1^{Y_{(1:2)}}\theta_2^{Y_{(2:2)}}\right) &= \theta_2\theta_1\overline{\boldsymbol{\pi}}_2\left(\mathbf{I}-\left(\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)\right)^{-1}\Big\{\left(\mathbf{t}_1\otimes\mathbf{t}_2\right)+\theta_2\left(\mathbf{T}_1\otimes\mathbf{t}_2\right)\left(\mathbf{I}-\left(\mathbf{T}_1\theta_2\right)\right)^{-1}\mathbf{t}_1\\
&\quad+\theta_2\left(\mathbf{t}_1\otimes\mathbf{T}_2\right)\left(\mathbf{I}-\left(\mathbf{T}_2\theta_2\right)\right)^{-1}\mathbf{t}_2\Big\}. \qquad (5.35)
\end{aligned}
$$

On the other hand, we are going to calculate the joint probability-generating function through Equation (5.6) and with representation

$$
\left(\left(\overline{\boldsymbol{\pi}}_2,\mathbf{0},\mathbf{0}\right),\mathbf{T}_{(2)};\mathbf{R}_{1,(2:2)}\right),
$$

where

$$
\mathbf{R}_{1,(2:2)} = \begin{array}{c} \begin{array}{cc} Y_{(1:2)} & Y_{(2:2)} \end{array}\\ \begin{pmatrix} \mathbf{e}_1\otimes\mathbf{e}_2 & \mathbf{e}_1\otimes\mathbf{e}_2\\ \mathbf{0} & \mathbf{e}_1\\ \mathbf{0} & \mathbf{e}_2 \end{pmatrix} \end{array},
$$

where $\mathbf{e}_i$ is the vector filled of ones of the same dimension than the initial vector $\boldsymbol{\alpha}_i, i = 1, 2$.

Then, the diagonal matrix of rewards is given by

$$
\Delta\left(\boldsymbol{\theta}^{\mathbf{r}}\right) = \begin{pmatrix} \mathbf{e}_1\otimes\mathbf{e}_2\theta_1\theta_2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{e}_1\theta_2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{e}_2\theta_2 \end{pmatrix}. \qquad (5.36)
$$

Now, we are going to calculate

$$
\left(\overline{\boldsymbol{\pi}}_2,\mathbf{0},\mathbf{0}\right)\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\left(\mathbf{I}-\mathbf{T}_{(2)}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\right)^{-1}\mathbf{t}_{(2)}.
$$

First, observe that

$$\left(\mathbf{I}-\mathbf{T}_{(2)}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\right)=\begin{pmatrix}(\mathbf{I}-\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2) & -\mathbf{T}_1\otimes\mathbf{t}_2\theta_2 & -\mathbf{t}_1\otimes\mathbf{T}_2\theta_2\\ \mathbf{0} & (\mathbf{I}-\mathbf{T}_1\theta_2) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & (\mathbf{I}-\mathbf{T}_2\theta_2)\end{pmatrix},$$

then

$$\left(\mathbf{I}-\mathbf{P}_{(2)}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\right)^{-1}=$$

$$\begin{pmatrix}\mathbf{C}_{1,2}(\theta_1\theta_2) & \theta_2\mathbf{C}_{1,2}(\theta_1\theta_2)\mathbf{B}_{1,2}\left(\mathbf{I}-\mathbf{T}_1\theta_2\right)^{-1} & \theta_2\mathbf{C}_{1,2}(\theta_1\theta_2)\mathbf{B}_{2,1}\left(\mathbf{I}-\mathbf{T}_2\theta_2\right)^{-1}\\ \mathbf{0} & \left(\mathbf{I}-\mathbf{T}_1\theta_2\right)^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \left(\mathbf{I}-\mathbf{T}_2\theta_2\right)^{-1}\end{pmatrix},$$

where $\mathbf{C}_{1,2}(\theta_1\theta_2)=\left(\mathbf{I}-\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{-1}$.

Thus,

$$\left(\overline{\boldsymbol{\pi}}_2,\mathbf{0},\mathbf{0}\right)\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\left(\mathbf{I}-\mathbf{T}_{(2)}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\right)^{-1}$$

$$=\begin{pmatrix}\overline{\boldsymbol{\pi}}_2\theta_1\theta_2\left(\mathbf{I}-\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{-1},\\ \overline{\boldsymbol{\pi}}_2\theta_1\theta_2^2\left(\mathbf{I}-\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{-1}\left(\mathbf{T}_1\otimes\mathbf{t}_2\right)\left(\mathbf{I}-\mathbf{T}_1\theta_2\right)^{-1},\\ \overline{\boldsymbol{\pi}}_2\theta_1\theta_2^2\left(\mathbf{I}-\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{-1}\left(\mathbf{t}_1\otimes\mathbf{T}_2\right)\left(\mathbf{I}-\mathbf{T}_2\theta_2\right)^{-1}\end{pmatrix}^{\top}.$$

Finally,

$$\begin{aligned}&\left(\overline{\boldsymbol{\pi}}_2,\mathbf{0},\mathbf{0}\right)\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\left(\mathbf{I}-\mathbf{T}_{(2)}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\right)^{-1}\mathbf{t}_{(2)}\\ =\;&\theta_1\theta_2\overline{\boldsymbol{\pi}}_2\left(\mathbf{I}-\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{-1}\left(\mathbf{t}_1\otimes\mathbf{t}_2\right)\\ +\;&\theta_1\theta_2^2\overline{\boldsymbol{\pi}}_2\left(\mathbf{I}-\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{-1}\left(\mathbf{T}_1\otimes\mathbf{t}_2\right)\left(\mathbf{I}-\mathbf{T}_1\theta_2\right)^{-1}\mathbf{t}_1\\ +\;&\theta_1\theta_2^2\overline{\boldsymbol{\pi}}_2\left(\mathbf{I}-\mathbf{T}_1\otimes\mathbf{T}_2\theta_1\theta_2\right)^{-1}\left(\mathbf{t}_1\otimes\mathbf{T}_2\right)\left(\mathbf{I}-\mathbf{T}_2\theta_2\right)^{-1}\mathbf{t}_2,\end{aligned}$$

which is exactly the expression for $\mathbb{E}\left(\theta_1^{Y_{(1:2)}}\theta_2^{Y_{(2:2)}}\right)$ in Equation (5.35).

Clearly, we can generalise the idea of the MPDH*-representation of simplest example for other cases. Let us consider the case of the bivariate $\left(Y_{(s:n)},Y_{(r:n)}\right)$, where $s<r$. Recall the representation given in Section 4.4 for the $r$-th order statistic. Thus, we have that the initial distribution is given by

$$\left(\boldsymbol{\pi}_1\otimes\cdots\otimes\boldsymbol{\pi}_n,\mathbf{0}\right)=\left(\overline{\boldsymbol{\pi}}_n,\mathbf{0}\right),$$

and the sub-transition probability matrix is

$$\mathbf{T}_{(r)} = \begin{pmatrix} \mathbf{C}_{(0,0)} & \mathbf{C}_{(0,1)} & \mathbf{C}_{(0,2)} & \cdots & \mathbf{C}_{(0,r-2)} & \mathbf{C}_{(0,r-1)} \\ \mathbf{0} & \mathbf{C}_{(1,1)} & \mathbf{C}_{(1,2)} & \cdots & \mathbf{C}_{(1,r-2)} & \mathbf{C}_{(1,r-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{(2,2)} & \cdots & \mathbf{C}_{(2,r-2)} & \mathbf{C}_{(2,r-1)} \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_{(r-2,r-2)} & \mathbf{C}_{(r-2,r-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{(r-1,r-1)} \end{pmatrix}.$$

Consider the matrix of rewards is

$$\mathbf{R}_{s,r:n} = \begin{array}{c} \begin{array}{cc} Y_{s:n} & Y_{r:n} \end{array} \\ \begin{pmatrix} \widehat{\mathbf{e}}_0 & \widehat{\mathbf{e}}_0 \\ \widehat{\mathbf{e}}_1 & \widehat{\mathbf{e}}_1 \\ \vdots & \vdots \\ \widehat{\mathbf{e}}_{s-1} & \widehat{\mathbf{e}}_{s-1} \\ \mathbf{0} & \widehat{\mathbf{e}}_s \\ \mathbf{0} & \widehat{\mathbf{e}}_{s+1} \\ \vdots & \vdots \\ \mathbf{0} & \widehat{\mathbf{e}}_{r-1} \end{pmatrix} \end{array},$$

where $\widehat{\mathbf{e}}_i$ is a column vector of ones of dimension equal to the number of columns of the matrix $\mathbf{C}_{(0,i)}, i = 0, 1, \ldots, r-1$.

Hence, $\left((\overline{\boldsymbol{\pi}}_n, \mathbf{0}), \mathbf{T}_{(r)}; \mathbf{R}_{s,r:n}\right)$ is a MDPH*-representation for the distribution of $(Y_{s:n}, Y_{r:n})$.

### 5.3.9 Compound multivariate phase-type distributions (CMPH)

Initially, one compound multivariate phase-type distributions was first introduced in the paper [HR16a] to model a multivariate insurance claim processes. Recently, a new discrete multivariate phase-type distributions was presented in [RZ17] based on a modification of the first compound multivariate phase-type distributions. This new discrete multivariate phase-type distribution generalises the discrete multivariate phase-type distribution defined in [HR16a] and, even more, it is considered as a modification of the discrete version of Kulkarni's multivariate phase-type distribution.

In the following, we are going to introduce the definition of the CMPH-distribution and explain how we can linked to the MDPH*-distributions.

Consider K categories of claims and assume that every accident occurs in a combination of those claims and every accident can be represented as an $K$-dimensional column vector

$$\mathbf{h} = (h_1, h_2, \ldots, h_K)^\top ,$$

where, for every $k = 1, \ldots, K$, $h_k$ is equal to 1 if the accident generates a claim of category $k$ and 0 otherwise. The set of all the possible column vectors, which represent the claims of the accidents, is denoted by $\mathcal{C}$. Accidents are assumed to be generated by a Markov chain. The compound multivariate phase-type distribution is defined next.

**Definition 5.14** Let $\{\mathcal{I}_m\}_{m\in\mathbb{N}}$ be a Markov chain with state space

$$\mathcal{E} = \{1, 2, \ldots, p, p+1\} ,$$

where the first $p$ states are transient and $p+1$ is the absorbing state.

Let $(\boldsymbol{\beta}, 1 - \boldsymbol{\beta}\mathbf{e})$ denote the initial distribution where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ encloses the initial probabilities of the first $p$ states, this is $\mathbb{P}(\mathcal{I}_0 = i) = \beta_i, i = 1, 2 \ldots, p$, respectively.

The transition probability matrix is given by

$$\begin{pmatrix} \mathbf{B} & \mathbf{b} \\ \mathbf{0} & 1 \end{pmatrix},$$

where $\mathbf{b} = \mathbf{e} - \mathbf{B}\mathbf{e}$.

It is assumed that any transition between the transient states can be accompanied by an accident $\mathbf{h} \in \mathcal{C}$, but transitions from transient states to the absorbing state are not accompanied by an accident and thus do not generate a batch.

Now, consider a decomposition of the matrix $\mathbf{B}$ as the sum

$$\mathbf{B} = \mathbf{B}_0 + \sum_{\mathbf{h}\in\mathcal{C}} \mathbf{B}_{\mathbf{h}},$$

where $\mathbf{B}_0$ is a sub-transition probability matrix which gives the probabilities of transitions between transient states that do not generate claims, this situation includes two cases: (1) when no accident occurs and consequently no claims and (2) when there is an accident but this accident does not generate claims; while $\mathbf{B}_{\mathbf{h}}$ denotes a sub-transition probability matrix where every transition is accompanied by an accident which generates claims.

For a given $\mathbf{h} \in \mathcal{C}$, let $\mathbf{U}_{\mathbf{h}} = (U_{\mathbf{h},1}, \ldots, U_{\mathbf{h},K})$, where $U_{\mathbf{h},k}$ is the size of the claim of category $k$ in $\mathbf{h}$.

Let $\mathbf{X}_{\mathbf{h}}$ denote the number of accidents $\mathbf{h}$ that occurred before the Markov chain $\{\mathcal{I}_m\}_{m\in\mathbb{N}}$ gets absorbed and $\mathbf{X}_{\mathbf{h}}$ is defined as follows.

(i) For $\mathbf{h} \in \mathcal{C}$, define $\xi_{\mathbf{h},0} = 0$ at time $0$.

(ii) If $\mathcal{I}_m = i \leq p$, then at time $m+1$.

(a) If $\mathcal{I}_{m+1} = j \leq p$ with probability $(\mathbf{B}_{\mathbf{h}})_{i,j}$, $\mathbf{h} \in \mathcal{C}$, then define $\xi_{\mathbf{h},m+1} = \xi_{\mathbf{h},m} + 1$ and $\xi_{\mathbf{u},m+1} = \xi_{\mathbf{u},m}$, for all $\mathbf{u} \neq \mathbf{h}$, $\mathbf{u} \in \mathcal{C}$.

(b) If $\mathcal{I}_{m+1} = j \leq p$ with probability $(\mathbf{B}_0)_{i,j}$, then define $\xi_{\mathbf{u},m+1} = \xi_{\mathbf{u},m}$, for all $\mathbf{u} \in \mathcal{C}$.

(c) If $\mathcal{I}_{m+1} = p+1$ with probability $1 - (\mathbf{Be})_i$, then define $\xi_{\mathbf{u},m+1} = \xi_{\mathbf{u},m}$, for all $\mathbf{u} \in \mathcal{C}$.

(iii) If $\mathcal{I}_m = p+1$, then the process is terminated and $\mathbf{X}_{\mathbf{h}} = \xi_{\mathbf{h},m}$ is defined for $\mathbf{h} \in \mathcal{C}$.

Now, for every $k$, which denotes the category of claim, consider the random variable

$$Y_k = \sum_{\mathbf{h} \in \mathcal{C}} \sum_{i=1}^{\mathbf{X}_{\mathbf{h}}} U_{\mathbf{h},k}(i), \quad k = 1, \ldots, K, \tag{5.37}$$

where $U_{\mathbf{h},k}(i), i \geq 1$, are independent copies of the random variable $U_{\mathbf{h},k}$.

Thus, $Y_k$ denote the total amount of losses produced by the claim of the category $k$ before the underlying Markov chain $\{\mathcal{I}_m\}_{m \in \mathbb{N}}$ gets absorbed. The distribution of the random vector $(Y_1, \ldots, Y_K)$ is called compound multivariate phase-type distribution (CMPH).

In [HR16a], the marginal variables of a discrete multivariate phase-type distribution is defined as

$$W_k = \sum_{\mathbf{h} \in \mathcal{C}} |\mathbf{h}|_k \mathbf{X}_{\mathbf{h}}, \quad 1 \leq k \leq K, \tag{5.38}$$

where $\mathbf{h}$ is a string of the numbers of the categories of the claims and a category can be presented more than one time. $|\mathbf{h}|_k$ is the number of times the claim of category $k$ appears in $\mathbf{h}$.

If we take

$$U_{\mathbf{h},k}(i) = |\mathbf{h}|_k, \quad \forall i \geq 1,$$

in Equation (5.37), then the variables $W_k$ coincides with the variable $Y_k$. In this way, the discrete multivariate phase-type distributions defined in [HR16a] are an particular case of the compound multivariate phase-type distributions.

Now, in order to explain the relation between CMPH-distributions and MDPH*-distributions, consider only one possible accident $\mathbf{h}$ in each transition and the case where $\mathbf{B}_0 = \mathbf{0}$ in the Markov chain $\{\mathcal{I}_m\}_{m\in\mathbb{N}}$. Then, the sub-transition probability matrix is reduced to

$$\mathbf{B} = \mathbf{B}_{\mathbf{h}},$$

and $\mathbf{X_h} = \tau - 1$. Consequently, the random variable $Y_k$ in Equation (5.37) becomes to

$$Y_k = \sum_{i=1}^{\tau-1} U_k(i),$$

where $U_k(i)$ is the size of the claim of category $k$ at time $i$, which can be also interpreted as the $k$-type reward obtained in the $i$-th transition of the Markov chain $\{\mathcal{I}_m\}_{m\in\mathbb{N}}$. Thus, if we denote $U_k(i) = r_k(\mathcal{I}_i)$, then the variable $Y_k$ can be written as

$$Y_k = \sum_{i=1}^{\tau-1} r_k(\mathcal{I}_i). \tag{5.39}$$

By Theorem 5.2, we know that $Y_k$ is discrete phase-type distributed and consequently $(Y_1, \ldots, Y_K)$ is MDPH*-distributed where the matrix of rewards $\mathbf{R}$ is derived from the $K$-types of rewards $(r_1, \ldots, r_K)$.

Observe that $Y_k$ in Equation (5.39) does not consider the reward at the time $0$. Therefore, the random vector $(Y_1, \ldots, Y_K)$ has a joint probability-generating function of the form

$$(1 - \boldsymbol{\beta}\mathbf{e}) + \boldsymbol{\beta}\left(\mathbf{I} - \mathbf{B}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\right)^{-1}\mathbf{b}.$$

## 5.4 Concluding remarks

In this chapter, we have introduced the discrete version of the multivariate phase-type distribution presented by Kulkarni in [Kul89]. We have shown a closed-form formula for the joint probability-generating function and, consequently, for the joint moment-generating function. As well as, we have proved some closure properties of this distribution and we have provided several examples of representations of it.

The formula of the joint probability-generating function is based on the first step analysis and the homogeneity of the underlying Markov chain. While the representations of the examples have been proposed from probabilistic interpretations and later have been verified analytically through the formula of the joint probability-generating function.

In Example 5.3.1, we remarked that to reduce the dimension of the representation and have a more convenient sub-transition probability matrix, it was more convenient to consider a different form for the joint probability-generating function which takes in account the reward of the absorbing state. The form of the formula in that case is

$$\boldsymbol{\pi}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\left(\mathbf{I}-\mathbf{T}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\right)^{-1}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\mathbf{t}.$$

In the same way, in Example 5.3.9 it would be more appropriate to consider another form of the joint probability-generating function which, in this case, does not take the rewards at time 0. Thus, in this case we may consider the formula

$$\boldsymbol{\pi}\left(\mathbf{I}-\mathbf{T}\boldsymbol{\Delta}\left(\boldsymbol{\theta}^{\mathbf{r}}\right)\right)^{-1}\mathbf{t}$$

for the joint probability-generating function.

Therefore, depending on the situation we are able to adapt the formula of the joint probability-generating function to the most convenient form. That shows the versatility of the approach given here for the study of the discrete version of multivariate phase-type distributions.

Lastly, notice that we did not work with the joint probability mass function of MDPH*-distributions. However, similarly to the continuous case, effective methods to compute these distributions are still missing in the literature due to numerical difficulties. However, we can find explicit formulas for particular cases, for instance in Example 5.3.5 we can find an explicit expression for the joint bivariate probability mass function in [MNOS10]. Recursive formulas have also been considered to calculate joint probability mass functions. That has been done by treating multivariate discrete phase-type distributions under different approaches, such as in Example 5.3.9 whose recursive formula is found in [RZ17].

# CHAPTER 6

# On concomitants of order statistics of phase-type distributions

In this chapter, we present our study on concomitants of order statistics from a sample of bivariate i.i.d. phase-type distributed random vectors, we show a procedure to calculate their distributions and prove that they are phase-type distributed. The base of the results come essentially from the following papers: A New Class Of Multivariate Phase-type Distributions [Kul89], A Stochastic Two-dimensional Fluid Model [BO13] and A Semi-explicit Density Function For Kulkarni's bivariate Phase-type Distributions [Bre16].

We start this chapter by explaining the concept of concomitants of order statistics and a couple of results on their formulas of distribution and density function. Then, we introduce the needed results given in [BO13] and [Bre16] in order to explain the contribution of this paper. After that, we present our study of concomitants from bivariate sample of i.i.d. phase-type distributed random vectors and the proof of concomitants are phase-type distributed. Lastly, we include a couple of examples of concomitants from different samples to clarify how to calculate their density function.

## 6.1 Introduction to concomitants of order statistics

Let $(X, Y)$ be a random vector with an absolutely continuous bivariate density function denoted by $h(x, y)$. Let $F_X(x)$ and $F_Y(y)$ denote the cumulative distribution function of $X$ and $Y$, respectively. For a fixed $n \in \mathbb{N}$, let us consider the sample

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_{n+1}, Y_{n+1}) \tag{6.1}$$

which are i.i.d pairs with density function $h(x, y)$. The size of the sample is kept to be $n + 1$ to ensure that the sample size is greater than one.

If the $Y$-variates in the bivariate sample (6.1) are arranged in increasing order, this yields to

$$\left(X_{[1:n+1]}, Y_{(1:n+1)}\right), \left(X_{[2:n+1]}, Y_{(2:n+1)}\right), \ldots, \left(X_{[n+1:n+1]}, Y_{(n+1:n+1)}\right),$$

then for every $r = 1, \ldots, n + 1$, the $X$-variate paired with the $r$-th order statistic $Y_{(r:n+1)}$ is called the **concomitant of the $r$-th order statistic** and it is denoted by $X_{[r:n+1]}$. In other way, based on the ordering of the $X$-variates, the concomitant of the $r$-th order statistic $X_{(r:n+1)}$ is defined as the $Y$-variate paired with $X_{(r:n+1)}$ and it is denoted by $Y_{[r:n+1]}$ for every $r = 1, \ldots, n + 1$.

**Proposition 6.1** *Consider the sample given in* (6.1) *and let $F_{r:n+1}(y)$ denote the cumulative distribution function of the $r$-th order statistic $Y_{(r:n+1)}$. Then, the cumulative distribution function of the concomitant of the $r$-th order statistic, $F_{[r:n+1]}(x)$, is given by*

$$F_{[r:n+1]}(x) = \int_{-\infty}^{\infty} \mathbb{P}\left(X \leq x | Y \in dy\right) dF_{r:n+1}(y). \tag{6.2}$$

**Proof.**

$$\begin{aligned}
\mathbb{P}\left(X_{[r:n+1]} \leq dx\right) &= \int_{-\infty}^{\infty} \mathbb{P}\left(X_{[r:n+1]} \leq dx, Y_{(r:n+1)} \in dy\right) dy \\
&= \int_{-\infty}^{\infty} \mathbb{P}\left(X_{[r:n+1]} \leq dx, Y_{(r:n+1)} \in dy, Y_i = Y_{(r:n+1)}\right) dy
\end{aligned} \tag{6.3}$$

for some $i \in \{1, \ldots, n + 1\}$. From here, the last expression satisfies

$$\begin{aligned}
&\int_{-\infty}^{\infty} \mathbb{P}\left(X_{[r:n+1]} \leq dx, Y_{(r:n+1)} \in dy, Y_i = Y_{(r:n+1)}\right) dy \\
= &\int_{-\infty}^{\infty} \mathbb{P}\left(X_{[r:n+1]} \leq dx, X_i = X_{[r:n+1]}, Y_i \in dy\right) dy
\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \mathbb{P}\left( X_{[r:n+1]} \leq dx \middle| X_i = X_{[r:n+1]}, Y_i \in dy \right) dF_{r:n+1}(y) \\
&= \int_{-\infty}^{\infty} \mathbb{P}\left( X_i \leq dx \middle| Y_i \in dy \right) dF_{r:n+1}(y).
\end{aligned}
$$

□

Notice that Equation (6.3) is only satisfied for continuous distributions. Due to that restriction, the distribution of the concomitants of order statistics in the discrete time has not been studied so far.

The next result shows the density function of concomitants, which is derived from Proposition 6.1.

**Proposition 6.2** *Let $f_{r:n+1}(y)$ denote the probability density function of the $r$-th order statistic $Y_{(r:n+1)}$ from the $Y$-variates of the sample* (6.1). *Then, the probability density function of the $r$-th concomitant, $f_{X_{[r:n+1]}}$, is given by*

$$
f_{X_{[r:n+1]}}(x) = r\binom{n+1}{r} \int_{-\infty}^{\infty} h(x,y) \left(F_Y(y)\right)^{r-1} \left(1 - F_Y(y)\right)^{n+1-r} dy.
$$

**Proof.** We recall that the density of the $r$-th order statistic $Y_{r:n+1}$ from i.i.d. random variables $Y_1, \ldots, Y_{n+1}$ with distribution function $F_Y(y)$ is calculated as follows

$$
f_{r:n+1}(y) = (n+1)\binom{n}{r-1} \left(F_Y(y)\right)^{r-1} \left(1 - F_Y(y)\right)^{n+1-r} f_Y(y), \tag{6.4}
$$

where $f_Y(y)$ denotes the density function of $Y$ (see [Bal07, p. 8]). Then, a formula for the density for the $r$-th concomitant $f_{X_{[r:n+1]}}(x)$ is derived next by combining Equations (6.2) and (6.4).

$$
\begin{aligned}
&f_{X_{[r:n+1]}}(x) \\
= \;& (n+1)\binom{n}{r-1} \int_{-\infty}^{\infty} \frac{d}{dx}\mathbb{P}\left(X \leq x | Y \in dy\right) f_Y(y) \left(F_Y(y)\right)^{r-1} \left(1 - F_Y(y)\right)^{n+1-r} dy \\
= \;& r\binom{n+1}{r} \int_{-\infty}^{\infty} h(x,y) \left(F_Y(y)\right)^{r-1} \left(1 - F_Y(y)\right)^{n+1-r} dy. \qquad (6.5)
\end{aligned}
$$

□

More properties of the distribution of concomitants such as their expectation, variance and covariance are found in [Yan77].

## 6.2 A stochastic two-dimensional fluid model.

In this section, we present the stochastic two-dimensional fluid model introduced by N. G. Bean and M. M. O'Reilly in [BO13]. Also, we present some results of this theory, such as the Laplace-Stieltjes transform of the distribution of a shift of the two dimensional fluid model, are going to be used to derive a semi-explicit formula for the distribution of a bivariate phase-type random variable.

**Definition 6.3 (Stochastic fluid model.)** A stochastic fluid model $\{(X_t, W_t)\}_{t\geq 0}$ is a process with parameters $\mathcal{E}, \mathbf{Q}$ and the set $\{q_i : \forall i \in \mathcal{E}\}$, where $\{X_t\}_{t\geq 0}$ is an irreducible Markov jump process with finite state space $\mathcal{E}$ and intensity matrix $\mathbf{Q}$, while $W_t$ is a continuous fluid level at time $t$ such that when $X_t = j$ it holds

$$\frac{dW_t}{dt} = \begin{cases} q_j & \text{if } W_t > 0, \\ \max\{0, q_j\} & \text{if } W_t = 0. \end{cases}$$

It says that the process $\{X_t\}_{t\geq 0}$ drives the evolution of $\{W_t\}_{t\geq 0}$.

Consider a stochastic fluid model $\{(X_t, W_t)\}_{t\geq 0}$ with a lower boundary $W_t \geq 0$, a finite and irreducible set of states $\mathcal{E}$ and intensity matrix $\mathbf{Q}$ with rates $s_i$ for all $i \in \mathcal{E}$ such that $s_i$ is a constant that can be positive, negative or zero. That is, $\{X_t\}_{t\geq 0}$ is a Markov jump process with state space $\mathcal{E}$ and intensity matrix $\mathbf{Q}$ and $\{X_t\}_{t\geq 0}$ drives the evolution of the level variable $W_t$ in the following way. When $X_t = i$ and $W_t > 0$, the rate at which $W_t$ is changing at time $t$ is $s_i$. However, when $X_t = i$ and $W_0 = 0$, the rate at which $W_t$ is changing at time $t$ is $\max\{0, r_i\}$ , and so if $s_i \leq 0$, there is no change in the level variable $W_t$ until a transition to some $j$, with $s_j > 0$, occurs.

Let $\{(X_t, Z_t, W_t)\}_{t\geq 0}$ be a stochastic model such that $\{X_t\}_{t\geq 0}$ is a Markov jump process with a finite and irreducible state space $\mathcal{E}$ and generator $\mathbf{Q}$. It is assumed that $\{(X_t, Z_t)\}_{t\geq 0}$ is an unbounded stochastic fluid model, this is $Z_t \in (-\infty, +\infty)$ , and rates $c_i$ such that for all $i \in \mathcal{E}$, $c_i$ is a constant that can be positive, negative or zero. When $X_t = i$, the rate at which $Z_t$ is changing at time $t$ is given by $c_i$. Next, it is assumed that $\{(X_t, W_t)\}_{t\geq 0}$ is a bounded stochastic fluid model with a lower boundary $W_t \geq 0$ and rates $s_i$, $i \in \mathcal{E}$, as it was defined earlier. The process $\{(X_t, Z_t, W_t)\}_{t\geq 0}$ is called a **stochastic two-dimensional fluid model** driven by $\{X_t\}_{t\geq 0}$ .

In the following we explain a method presented in [BO13] that considers paths of the process $\{(X_t, W_t)\}_{t\geq 0}$ under which we can evaluate the corresponding distribution of $Z_t - Z_0$ (referred to as the shift) in the process $\{(X_t, Z_t)\}_{t\geq 0}$. The Laplace-Stieltjes transform of the distribution of the shift $Z_t - Z_0$ corresponding to a sample path in $\{(X_t, W_t)\}_{t\geq 0}$ is going to be expressed in terms of matrices.

Consider the process $\{(X_t, W_t)\}_{t\geq 0}$. Let $\mathcal{E}_1 = \{i \in \mathcal{E} : s_i > 0\}$ , $\mathcal{E}_2 = \{i \in \mathcal{E} : s_i < 0\}$ and $\mathcal{E}_0 = \{i \in \mathcal{E} : s_i = 0\}$ . The intensity matrix $\mathbf{Q}$ is partitioned according to the subsets of $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_0$, as follows

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} & \mathbf{Q}_{10} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} & \mathbf{Q}_{20} \\ \mathbf{Q}_{01} & \mathbf{Q}_{02} & \mathbf{Q}_{00} \end{pmatrix}.$$

Let $f(t) = \int_0^t \left| s_{X_u} \right| du$. From here, notice that $f(t) \geq 0$ and it can be interpreted as the total amount of fluid that has flowed into or out of the $W_{(\cdot)}$ during the time interval $(0, t)$.

Assume that $W_0 = 0$. Let $\omega(y) = \inf \{t > 0 : f(t) = y\}$ be the first time the total amount of fluid that has flowed into or out of $W_{(\cdot)}$ reaches the level $y$.

For $i, j \in \mathcal{E}_1 \cup \mathcal{E}_2$, let

$$\delta_i^y(j, t) = \mathbb{P}\left(\omega(y) \leq t, X_{\omega(y)} = j \,\middle|\, W_0 = 0, X_0 = i\right). \tag{6.6}$$

The function $\delta_i^y(j, t)$ is the joint distribution function/probability mass that by starting from level zero in state $i$, the total amount of fluid that has flowed into or out of $W_{(\cdot)}$ first reaches $y$ at time less than or equal to $t$, and it does so in state $j$. Let $\boldsymbol{\delta}^y(u)$ be a matrix such that $[\boldsymbol{\delta}^y(u)]_{i,j} = \delta_i^y(j, u)$.

Now, let us consider the process $\{(X_t, Z_t)\}_{t \geq 0}$. The aim is to derive the distribution of the shift $Z_t - Z_0$. For that purpose, let us consider the following Laplace-Stieltjes transform.

Let $\boldsymbol{\Delta}^y(s)$ be the matrix whose $(i, j)$-th entry is given by

$$[\boldsymbol{\Delta}^y(s)]_{i,j} = \int_0^\infty e^{-sS_t} d\delta_i^y(j, t),$$

where $S_t = Z_t - Z_0$.

In what follows, we analyse some conditions for $\boldsymbol{\Delta}^y(s)$ to be well-defined. Once the conditions are established, $[\boldsymbol{\Delta}^y(s)]_{i,j}$ is going to be interpreted as the Laplace-Stieltjes transform of the distribution of $Z_t - Z_0$ at time $\omega(y)$, assuming $X_0 = i$, $W_0 = 0$ and $X_{\omega(y)} = j$.

Let us introduce the following matrices.

$$\mathbf{R}_1 = \text{diag}\,(s_i)_{i \in \mathcal{E}_1} \qquad \text{and} \qquad \mathbf{R}_2 = \text{diag}\,(|s_i|)_{i \in \mathcal{E}_2}.$$

Also, let

$$\mathbf{D}_0 = \text{diag}\,(c_i)_{i \in \mathcal{E}_0}, \quad \mathbf{D}_1 = \text{diag}\,(c_i)_{i \in \mathcal{E}_1} \quad \text{and} \quad \mathbf{D}_2 = \text{diag}\,(c_i)_{i \in \mathcal{E}_2}.$$

For $i, j \in \mathcal{E}_0$, it is defined the joint distribution function/probability mass function

$$\beta_i^t(j, x) = \mathbb{P}\left(S_t \leq x, X_t = j \,|\, X_0 = i, X_u \in \mathcal{E}_0, 0 \leq u \leq t\right). \tag{6.7}$$

Let $\boldsymbol{\beta}^t(x)$ be a matrix such that $\left[\boldsymbol{\beta}^t(x)\right]_{i,j} = \beta_i^t(j, x)$.

Let $\mathbf{B}^t(s)$ be the matrix of the two-sided Laplace-Stieltjes transforms whose $(i, j)$-th entry is

$$\left[\mathbf{B}^t(s)\right]_{ij} = \int_{-\infty}^{+\infty} e^{-sx} d\beta_i^t(j, x). \tag{6.8}$$

Then, $\left[\mathbf{B}^t(s)\right]_{ij}$ is interpreted as the Laplace-Stieltjes transform of the distribution of the shift in $S_{(.)}$ over the interval $(0, t)$, with $X_t = j$, assuming $X_0 = i$, and $X_u \in \mathcal{E}_0, 0 \leq u \leq t$.

**Theorem 6.4** *For every $s \in \mathbb{C}$, the matrix $\mathbf{B}^t(s)$ is well-defined for all $t \geq 0$ and is given by*

$$\mathbf{B}^t(s) = \exp\left\{\left(\mathbf{Q}_{00} - s\mathbf{D}_0\right)t\right\}. \tag{6.9}$$

For the proof we refer [BO13, Theorem 1].

Let $\mathbf{A}$ be a matrix and let $\chi(\mathbf{A})$ denote the eigenvalue with maximum real part of $\mathbf{A}$.

**Theorem 6.5** *If $s$ is such that*

$$\chi\left(\mathbf{Q}_{00} - s\mathbf{D}_0\right) < 0, \tag{6.10}$$

*then $\boldsymbol{\Delta}^y(s)$ exists and is given by*

$$\boldsymbol{\Delta}^y(s) = \exp\left\{\mathbf{W}(s)y\right\}, \tag{6.11}$$

*where*

$$\mathbf{W}(s) = \begin{pmatrix} \mathcal{W}_{11}(s) & \mathcal{W}_{12}(s) \\ \mathcal{W}_{21}(s) & \mathcal{W}_{22}(s) \end{pmatrix},$$

*with*

$$\begin{aligned}
\mathcal{W}_{11}(s) &= \mathbf{R}_1^{-1}\left[\left(\mathbf{Q}_{11} - s\mathbf{D}_1\right) - \mathbf{Q}_{10}\left(\mathbf{Q}_{00} - s\mathbf{D}_0\right)^{-1}\mathbf{Q}_{01}\right], \\
\mathcal{W}_{22}(s) &= \mathbf{R}_2^{-1}\left[\left(\mathbf{Q}_{22} - s\mathbf{D}_2\right) - \mathbf{Q}_{20}\left(\mathbf{Q}_{00} - s\mathbf{D}_0\right)^{-1}\mathbf{Q}_{02}\right], \\
\mathcal{W}_{12}(s) &= \mathbf{R}_1^{-1}\left[\mathbf{Q}_{12} - \mathbf{Q}_{10}\left(\mathbf{Q}_{00} - s\mathbf{D}_0\right)^{-1}\mathbf{Q}_{02}\right], \\
\mathcal{W}_{21}(s) &= \mathbf{R}_1^{-1}\left[\mathbf{Q}_{21} - \mathbf{Q}_{20}\left(\mathbf{Q}_{00} - s\mathbf{D}_0\right)^{-1}\mathbf{Q}_{01}\right].
\end{aligned}$$

For the proof we refer [BO13, Theorem 2].

**Lemma 6.6** *The set* $\mathscr{W} = \{s \in \mathbb{C} : \chi(\mathbf{Q}_{00} - s\mathbf{D}_0) < 0\}$ *is a non-trivial, unbounded vertical strip in the complex plane.*

See proof in [BO13, Lemma 2].

**Lemma 6.7** *The set* $\mathscr{P} = \{s \in \mathscr{W} : \chi(\mathcal{W}_{11}(s)) < 0$ *and* $\chi(\mathcal{W}_{22}(s)) < 0\}$ *is a non-trivial, unbounded vertical strip in the complex plane.*

See proof in [BO13, Lemma 3].

## 6.3 The class of MPH*

In the introduction of the thesis we talked for the first time about the class of MPH*, however we did so without substantial details. Here, we present formally the definition of the class of multivariate phase-type distributions introduced by V. G. Kulkarni in [Kul89] and remark important assumptions. Furthermore, we recall the join moment-generating function of MPH* and the closure property of its marginals. Lastly, to finalise this section, we present an easy example of a bivariate phase-type distribution.

Let $\{X_t\}_{t\geq 0}$ be a Markov jump process with state space

$$\mathcal{E} = \{1, 2, \ldots, p, p+1\},$$

where the first $p$ states are transient and $p+1$ is an absorbing state.

The intensity matrix can be written as follows

$$\begin{pmatrix} \mathbf{S} & -\mathbf{S}\mathbf{e} \\ \mathbf{0} & 0 \end{pmatrix}, \tag{6.12}$$

where $\mathbf{S}$ is an $p$-square matrix, $\mathbf{e}$ is an $p$-dimensional column vector filled of one and $\mathbf{0}$ is an $p$-dimensional row zero vector.

Furthermore, let $(\boldsymbol{\alpha}, \alpha_{p+1})$ denote the initial distribution of $\{X_t\}_{t\geq 0}$ where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$.

Consider the first passage time of the Markov jump process

$$\tau = \inf\{t \geq 0 : X_t = p+1\}. \tag{6.13}$$

Let $\mathbf{r} = (r(1), \ldots, r(p))^\top$ be an $p$-dimensional column vector with non-negative entries. Every entry $r(i)$ represents the reward obtained in state $i$. We refer to $\mathbf{r}$ as "the vector of rewards".

Consider the variable

$$\mathcal{Y} = \int_0^\tau r(X_t)dt. \tag{6.14}$$

Then, $\mathcal{Y}$ represents the total reward accumulated until the process $\{X_t\}_{t\geq 0}$ gets absorbed. Notice that if $r(i) = 1$ for all $i = 1, \ldots, p$, then $\mathcal{Y} = \tau$.

**Theorem 6.8** *The random variable $\mathcal{Y}$ is phase-type distributed.*

See [Kul89] for the proof.

Now, this idea is generalised by introducing more vector of rewards. Consider $n$ vectors of rewards denoted by

$$\mathbf{r}_k = (r_k(1), \ldots, r_k(p))^\top, \quad k = 1, \ldots, n. \tag{6.15}$$

These vectors form a matrix defined as follows

$$\mathbf{R} = \begin{pmatrix} \mathbf{r}_1 & \cdots & \mathbf{r}_n \end{pmatrix},$$

which is of dimension $p \times n$. The matrix $\mathbf{R}$ is referred as "the matrix of rewards".

For every vector of rewards consider the variable

$$\mathcal{Y}_k = \int_0^\tau r_k(X_t)dt, \quad k = 1, \ldots, n. \tag{6.16}$$

Then, the random vector $(\mathcal{Y}_1, \ldots, \mathcal{Y}_n)$ is said to be multivariate phase-type distributed (MPH$^*$-distributed). We write $(\mathcal{Y}_1, \ldots, \mathcal{Y}_n) \sim \mathrm{MPH}_p^* (\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R})$, where $p$ denotes the dimension of the PH-representation $(\boldsymbol{\alpha}, \mathbf{S})$.

**Theorem 6.9 (Joint moment-generating function.)** *Let $(\mathcal{Y}_1, \ldots, \mathcal{Y}_n) \sim \mathrm{MPH}_p^* (\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R})$. Then, there exists a constant $\theta_0 > 0$ such that the joint moment-generating function*

$$\mathbb{E}\left(\exp\{-(\theta_1\mathcal{Y}_1 + \cdots + \theta_n\mathcal{Y}_n)\}\right)$$

*exists for any $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ with $\theta_i < \theta_0$, and it is given by*

$$\mathbb{E}\left(\exp\{-(\theta_1\mathcal{Y}_1 + \cdots + \theta_n\mathcal{Y}_n)\}\right) = \boldsymbol{\alpha}\left(-\boldsymbol{\Delta}\left(\mathbf{R}\boldsymbol{\theta}\right) - \mathbf{S}\right)^{-1}\mathbf{s}, \tag{6.17}$$

*where $\mathbf{s} = -\mathbf{S}\mathbf{e}$ and $\boldsymbol{\Delta}(\mathbf{R}\boldsymbol{\theta})$ is the diagonal matrix formed with the entries of the vector*

$$\mathbf{R}\boldsymbol{\theta} = \left(\sum_{k=1}^{n} r_k(1)\theta_1, \ldots, \sum_{k=1}^{n} r_k(p)\theta_1\right).$$

For the proof we refer [BN17, p. 438].

**Proposition 6.10** *If* $(\mathcal{Y}_1, \dots, \mathcal{Y}_n)$ *is* MPH$^*$*-distributed, then every variable* $\mathcal{Y}_i$ *is phase-type distributed for* $i = 1, \dots, n$.

The proof is given in [Kul89].

**Remark.** Throughout this chapter, we assume that

$$\sum_{i=1}^{p} r_k(i) > 0,$$

for every fixed vector of rewards $\mathbf{r}_k, k = 1, \dots, n$. Hence, $\mathbb{P}(\mathcal{Y}_k > 0) > 0$, for all $k = 1, \dots, n$.

Furthermore, we assume (without loss of generality) that

$$\sum_{k=1}^{n} r_k(i) > 0,$$

for every fixed state $i \in \mathcal{E}$. That is assumed to avoid superfluous states.

### 6.3.1 Example: a bivariate distribution of two Erlang distributed random variables.

Let $Z$, $W_1$ and $W_2$ be independent and identical exponentially distributed random variables with intensity 1. Consider the variables

$$\mathcal{Y}_i = \frac{Z + W_i}{\lambda_i}, \quad i = 1, 2.$$

Then, $\mathcal{Y}_i$ is Erlang distributed with parameters $(2, \lambda_i)$ for every $i = 1, 2$, (see Example 3.2). A MPH$^*$-representation for the distribution of $(\mathcal{Y}_1, \mathcal{Y}_2)$ is $(\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R})$, where

$$\boldsymbol{\alpha} = (1, 0, 0), \quad \mathbf{S} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}, \quad \mathbf{R} = \begin{array}{c} \begin{array}{cc} \mathcal{Y}_1 & \mathcal{Y}_2 \end{array} \\ \begin{pmatrix} \frac{1}{\lambda_1} & \frac{1}{\lambda_2} \\ \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{pmatrix} \end{array}.$$

We can verify that $(\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R})$ is a MPH$^*$-representation for $(\mathcal{Y}_1, \mathcal{Y}_2)$ by calculating its joint moment-generating function directly as follows.

$$\mathbb{E}\left(\exp\left\{-\theta_1 \mathcal{Y}_1 - \theta_2 \mathcal{Y}_2\right\}\right)$$

$$
\begin{aligned}
&= \mathbb{E}\left(\exp\left\{-\left(\frac{\theta_1}{\lambda_1}+\frac{\theta_2}{\lambda_2}\right)Z\right\}\right)\mathbb{E}\left(\exp\left\{-\frac{\theta_1}{\lambda_1}W_1\right\}\right)\mathbb{E}\left(\exp\left\{-\frac{\theta_1}{\lambda_1}W_1\right\}\right)\\
&= \left(1-\left(\frac{\theta_1}{\lambda_1}+\frac{\theta_2}{\lambda_2}\right)\right)^{-1}\left(1-\frac{\theta_1}{\lambda_1}\right)^{-1}\left(1-\frac{\theta_2}{\lambda_2}\right)^{-1}\\
&= \frac{(\lambda_1\lambda_2)^2}{(\lambda_1\lambda_2-\theta_1\lambda_2-\theta_2\lambda_1)(\lambda_1-\theta_1)(\lambda_2-\theta_2)}\\
&= \boldsymbol{\alpha}\left(-\boldsymbol{\Delta}\left(\mathbf{R}\,\boldsymbol{\theta}\right)-\mathbf{S}\right)^{-1}\mathbf{s},
\end{aligned}
$$

where the last equality is verified by making the calculation of the formula (6.17) in Theorem 6.9.

## 6.4 A semi-explicit density function for Kulkarni's bivariate phase-type distributions

In this section, our goal is to present the formula for the semi-explicit Laplace transform of a bivariate phase-type distribution derived by L. Breuer in [Bre16].

Consider the bivariate random vector $(\mathcal{Y}_1, \mathcal{Y}_2) \sim \text{MPH}^*(\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R})$ with underlying Markov jump process $\{X_t\}_{t\geq 0}$ and state space $\mathcal{E} = \{1, 2, \dots, p, p+1\}$.

In [Bre12] is shown that the marginal $\mathcal{Y}_1$ holds a relation with the first passage time of a stochastic fluid process and that relation is actually an essential key in the proof of the semi-explicit formula for the density function. We explain that next.

Consider the variable defined by

$$W_t := \int_0^t r_1(X_s)ds, \tag{6.18}$$

for $t \geq 0$, where $r_1(\cdot)$ denotes the rewards comming from the random variable $\mathcal{Y}_1$ and $r_1(p+1) := 0$. Then, the multivariate process $\{(X_t, W_t)\}_{t\geq 0}$ is a stochastic fluid process.

Observe that the process $\{W_t\}_{t\geq 0}$ does not have decreasing paths due to $r_1(i) \geq 0$ for every state $i \in \mathcal{E}$. Then, the equalities

$$W_t = W_\tau = \mathcal{Y}_1 \tag{6.19}$$

hold for all $t \geq \tau$.

Consider the first passage times given by

$$T(y) := \inf\{t \geq 0 : W_t > y\}, \quad y \geq 0, \tag{6.20}$$

where $\inf \emptyset := \infty$. Then, it holds that

$$\mathbb{P}(\mathcal{Y}_1 > y) = \mathbb{P}_{\boldsymbol{\alpha}}(T(y) < \tau), \tag{6.21}$$

where $\tau \sim \mathrm{PH}(\boldsymbol{\alpha}, \mathbf{S})$ and $\mathbb{P}_{\boldsymbol{\alpha}}$ denotes the conditional probability given $\mathbb{P}(X_0 = i) = \alpha_i, i \in \mathcal{E}$.

**Remark.** From the definition of the first passage times (see Equation (6.20)), notice that the event $\{T(y) \geq \tau\}$ implies that there is a positive reward on the state $p+1$, which is impossible. Therefore, the event $\{T(y) < \infty\}$ is equivalent to the event $\{T(y) < \tau\}$. From here, we have that

$$\mathbb{P}(\mathcal{Y}_1 > y) = \mathbb{P}_{\boldsymbol{\alpha}}(T(y) < \infty) \tag{6.22}$$

for all $y \geq 0$.

Now consider the stochastic fluid process given by $\{(X_t, Z_t)\}_{t \geq 0}$, where

$$Z_t := \int_0^t r_2(X_s)ds, \tag{6.23}$$

for all $t \geq 0$, and $r_2(p+1) := 0$.

This definition is completely analogous to the definition of $\{W_t\}_{t\geq 0}$ above (see Equation (6.18)). Thus, the paths of the process $\{Z_t\}_{t\geq 0}$ are also non-decreasing paths and it is satisfied that

$$Z_t = Z_\tau = \mathcal{Y}_2 \tag{6.24}$$

for all $t \geq \tau$.

Consider the partition of the state space $\mathcal{E}$ given by $\mathcal{E} = \mathcal{E}_0 \cup \mathcal{E}_+$, where

$$\mathcal{E}_0 := \{i \in \mathcal{E} : r_1(i) = 0\} \quad \text{and} \quad \mathcal{E}_+ := \{i \in \mathcal{E} : r_1(i) > 0\}.$$

According to this partition, we can write the sub-intensity matrix $\mathbf{S}$ and the initial vector $\boldsymbol{\alpha}$ in the following block form

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{00} & \mathbf{S}_{0+} \\ \mathbf{S}_{+0} & \mathbf{S}_{++} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\alpha} = (\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_+).$$

Also, we write the exit vector as $-\mathbf{S}\mathbf{e} = (\mathbf{s}_0, \mathbf{s}_+)^\top$.

$\mathbf{S}_{00}$ is the sub-intensity matrix formed with the transition rates between the states in $\mathcal{E}_0$. Then, $\mathbf{S}_{0+}$ is the sub-intensity matrix consisted by with the transition rates from states in $\mathcal{E}_0$ to states in $\mathcal{E}_+$. Similarly, In the other way, $\mathbf{S}_{+0}$ is the sub-intensity matrix formed with the transition rates from states in $\mathcal{E}_+$ to states in $\mathcal{E}_0$. Lastly, $\mathbf{S}_{++}$ is the sub-intensity matrix formed with the transition rates between the states in $\mathcal{E}_+$. In the initial probability vector, $\boldsymbol{\alpha}_0$ is the row vector of initial probabilities of the states in $\mathcal{E}_0$ and $\boldsymbol{\alpha}_+$ is the row vector of initial probabilities of the states in $\mathcal{E}_+$. While in the exit vector, $\mathbf{s}_0$ is the column vector of the exit rates of the states in $\mathcal{E}_0$ and $\mathbf{s}_+$ is the column vector of the exit rates of the states in $\mathcal{E}_+$.

Finally, consider the diagonal matrices

$$\mathbf{R}_+ := \text{diag}\left(r_1(i) : i \in \mathcal{E}_+\right), \quad \mathbf{R}_0 := \mathbf{0},$$
$$\mathbf{D}_+ := \text{diag}\left(r_2(i) : i \in \mathcal{E}_+\right), \quad \text{and} \quad \mathbf{D}_0 := \text{diag}\left(r_2(i) : i \in \mathcal{E}_0\right).$$

**Note.** Since all states are assumed to be not superfluous, we have that $r_2(i) > 0$ for all $i \in \mathcal{E}_0$, and this implies that $\mathbf{D}_0$ is nonsingular.

Also notice that if $\mathcal{E}_0 \neq \emptyset$, then $\mathcal{Y}_1$ has an atom at zero. The next results shows an expression for the density function of the marginal $\mathcal{Y}_2$ under the event $\{\mathcal{Y}_1 = 0\}$.

**Theorem 6.11** *On the set $\{\mathcal{Y}_1 = 0\}$, $\mathcal{Y}_2$ has a deficient phase-type distribution with density function*

$$\mathbb{P}\left(\mathcal{Y}_2 \in dx, \mathcal{Y}_1 = 0\right) = \boldsymbol{\alpha}_0 \exp\left(\mathbf{D}_0^{-1}\,\mathbf{S}_{00}\,x\right)\mathbf{D}_0^{-1}\mathbf{s}_0\,dx,$$

*for all $x > 0$.*

**Proof.** The event $\{\mathcal{Y}_1 = 0\}$ happens if and only if $X_s \in \mathcal{E}_0$ for all $s \leq \tau$, this means that $\{X_t\}_{t\geq 0}$ never enters to $\mathcal{E}_+$. Consequently, from the relation

$$\mathcal{Y}_2 = Z_\tau = \int_0^\tau r_2(X_s)ds$$

on the event $\{\mathcal{Y}_1 = 0\}$, we conclude that $\mathcal{Y}_2$ has a deficient phase-type distribution with initial phase distribution $\boldsymbol{\alpha}_0$, sub-intensity matrix $\mathbf{D}_0^{-1}\mathbf{S}_{00}$ and exit vector

$$-\mathbf{D}_0^{-1}\left(\mathbf{S}_{00} + \mathbf{S}_{0+}\right)\mathbf{e} = \mathbf{D}_0^{-1}\mathbf{s}_0.$$

Also, notice that the exit vector does not count the exit rates of the sub-intensity matrix $\mathbf{S}_{0+}$, because (again), under the set $\{\mathcal{Y}_1 = 0\}$, $\{X_t\}_{t\geq 0}$ never enters to $\mathcal{E}_+$. □

Let $\mathbb{E}\left(e^{-s\mathcal{Y}_2}\mathbf{1}_{\{\mathcal{Y}_1=0\}}\right) =: \mathbb{E}\left(e^{-s\mathcal{Y}_2}, \mathcal{Y}_1 = 0\right)$. Then, the Laplace transform of the distribution of $\mathcal{Y}_2$ on $\{\mathcal{Y}_1 = 0\}$ is calculated as follows

$$\mathbb{E}\left(e^{-s\mathcal{Y}_2}, \mathcal{Y}_1 = 0\right) = \int_0^\infty e^{-sx}\boldsymbol{\alpha}_0 \exp\left\{\mathbf{D}_0^{-1}\mathbf{S}_{00}\,x\right\}\mathbf{D}_0^{-1}\mathbf{s}_0\,dx$$

$$
\begin{aligned}
&= \boldsymbol{\alpha}_0 \int_0^\infty \exp\left\{\left(\mathbf{D}_0^{-1}\mathbf{S}_{00} - s\mathbf{I}\right)x\right\} dx \mathbf{D}_0^{-1}\mathbf{s}_0 \\
&= -\boldsymbol{\alpha}_0 \left(\mathbf{D}_0^{-1}\mathbf{S}_{00} - s\mathbf{I}\right)^{-1} \mathbf{D}_0^{-1}\mathbf{s}_0, \quad s \geq 0, \text{(see Equation (A.3))} \\
&= \boldsymbol{\alpha}_0 \left(s\mathbf{D}_0 - \mathbf{S}_{00}\right)^{-1} \mathbf{s}_0. \qquad (6.25)
\end{aligned}
$$

The next result covers the case when $\mathcal{Y}_1 > 0$.

Let $\mathbb{E}\left(e^{-s\mathcal{Y}_2}\mathbf{1}_{\{\mathcal{Y}_1 \in dy\}}\right) =: \mathbb{E}\left(e^{-s\mathcal{Y}_2}, \mathcal{Y}_1 \in dy\right)$.

**Theorem 6.12** *Let* $(\mathcal{Y}_1, \mathcal{Y}_2) \sim \text{MPH}^*\left(\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R}\right)$. *Then*

$$
\mathbb{E}\left(e^{-s\mathcal{Y}_2}, \mathcal{Y}_1 \in dy\right) = \alpha(s)\, e^{\mathcal{W}(s)y}\eta(s)dy, \qquad (6.26)
$$

*for* $y > 0$ *and* $s \geq 0$, *where*

$$
\begin{aligned}
\alpha(s) &:= \boldsymbol{\alpha}_0\Big(s\mathbf{D}_0 - \mathbf{S}_{00}\Big)^{-1}\mathbf{S}_{0+} + \boldsymbol{\alpha}_+ \\
\mathcal{W}(s) &:= \mathbf{R}_+^{-1}\Big(\left(\mathbf{S}_{++} - s\mathbf{D}_+\right) - \mathbf{S}_{+0}\left(\mathbf{S}_{00} - s\mathbf{D}_0\right)^{-1}\mathbf{S}_{0+}\Big) \\
\eta(s) &:= \mathbf{R}_+^{-1}\Big(\mathbf{S}_{+0}\left(s\mathbf{D}_0 - \mathbf{S}_{00}\right)^{-1}\mathbf{s}_0 + \mathbf{s}_+\Big).
\end{aligned}
$$

**Proof.** The proof is originally given in [Bre16]. However, in the following we make an outline of the proof.

Let

$$
\mathbb{E}\left(\exp\left\{-s\int_0^{T(y)} r_2\left(X_u\right)du\right\}, T(y) < \tau\right)
$$

denote the matrix with $(i, j)$-th entry given by

$$
\mathbb{E}\left(\exp\left\{-s\int_0^{T(y)} r_2\left(X_u\right)du\right\}, T(y) < \tau, X_{T(y)} = j \,\middle|\, X_0 = i\right) \qquad (6.27)
$$

for $i, j \in \mathcal{E}_+$.

Recall Theorem 6.5 and also notice that in our case we have $S_{T(y)} = \displaystyle\int_0^{T(y)} r_2\left(X_u\right)du$ and

$$
\mathbb{E}\left(\exp\left\{-s\int_0^{T(y)} r_2\left(X_u\right)du\right\}, T(y) < \tau, X_{T(y)} = j \,\middle|\, X_0 = i\right)
$$

$$= \quad \Big[ \exp \{ \mathcal{W}(s) y \} \Big]_{i,j}$$

for $i, j \in \mathcal{E}_+$.

Now, since events $\{T(y) < \tau\}$ and $\{\mathcal{Y}_1 > y\}$ are equivalent (see Equation (6.21)), it holds that

$$\begin{aligned} & \mathbb{E}\left( \exp\left\{ -s \int_0^{T(y)} r_2\left(X_u\right) du \right\}, T(y) < \tau, X_{T(y)} = j \middle| X_0 = i \right) \\ = \; & \mathbb{E}\left( \exp\left\{ -s \int_0^{T(y)} r_2\left(X_u\right) du \right\}, \mathcal{Y}_1 > y, X_{T(y)} = j \middle| X_0 = i \right) \end{aligned}$$

and consequently

$$\begin{aligned} & \mathbb{E}\left( \exp\left\{ -s \int_0^{T(y)} r_2\left(X_u\right) du \right\}, \mathcal{Y}_1 > y, X_{T(y)} = j \middle| X_0 = i \right), \\ = \; & \Big[ \exp \{ \mathcal{W}(s) y \} \Big]_{i,j} \end{aligned}$$

for $i, j \in \mathcal{E}_+$.

Next, the goal is to find an expression for $\mathbb{E}\left( e^{-s\mathcal{Y}_2}, \mathcal{Y}_1 \in dy \middle| X_0 = i \right)$ and for that this expectations is written as follows

$$\mathbb{E}\left( e^{-s\mathcal{Y}_2}, \mathcal{Y}_1 \in dy \middle| X_0 = i \right) = \lim_{h \downarrow 0} \frac{1}{h} \mathbf{e}_i^\top \mathbb{E}\left( e^{-s\mathcal{Y}_2}, y < \mathcal{Y}_1 \le y + h \right),$$

where $i \in \mathcal{E}_+$ and $\mathbb{E}\left( e^{-s\mathcal{Y}_2}, y < \mathcal{Y}_1 \le y + h \right)$ denotes the column vector of dimension $|\mathcal{E}_+|$ with entries $\mathbb{E}\left( e^{-s\mathcal{Y}_2}, y < \mathcal{Y}_1 \le y + h \middle| X_0 = i \right)$.

From the strong Markov property, it follows that

$$\begin{aligned} \mathbb{E}\left( e^{-s\mathcal{Y}_2}, y < \mathcal{Y}_1 \le y + h \right) &= \mathbb{E}\left( \exp\left\{ -s \int_0^{\tau} r_2\left(X_u\right) du \right\}, y < \mathcal{Y}_1 \le y + h \right) \\ &= \mathbb{E}\left( \exp\left\{ -s \int_0^{T(y)} r_2\left(X_u\right) du \right\}, \mathcal{Y}_1 > y \right) \\ &\times \; \mathbb{E}\left( \exp\left\{ -s \int_{T(y)}^{\tau} r_2\left(X_u\right) du \right\}, \mathcal{Y}_1 \le y + h \middle| \mathcal{Y}_1 > y \right), \end{aligned}$$

where

$$\mathbb{E}\left( \exp\left\{ -s \int_{T(y)}^{\tau} r_2\left(X_u\right) du \right\}, \mathcal{Y}_1 \le y + h \middle| \mathcal{Y}_1 > y \right)$$

is a column vector of dimension $|\mathcal{E}_+|$ with entries

$$\mathbb{E}\left( \exp\left\{ -s \int_{T(y)}^{\tau} r_2\left(X_u\right) du \right\}, \mathcal{Y}_1 \le y + h \middle| \mathcal{Y}_1 > y, X_{T(y)} = i \right)$$

for $i \in \mathcal{E}_+$.

By now it is assumed that $X_{T(y)} = j$, where $j \in \mathcal{E}_+$. Consider the paths from which $\{X_t\}_{\geq 0}$ leaves the set $\mathcal{E}_+$ within the interval $[T(y), T(y) + (h/r_1(j))]$ . Thus, $\{X_t\}_{\geq 0}$ have two types of paths which are explained next.

The first type of paths is given when $\{X_t\}_{\geq 0}$ goes from state $j$ to the absorbing state $p+1$, which happens with probability $(h\mathbf{s}_+(j))/r_1(j)$, where $\mathbf{s}_+(j)$ denotes the $j$-th entry of the vector $\mathbf{s}_+$, and in this case the contribution of

$$\int_{T(y)}^{\tau} r_2\left(X_u\right) du$$

will vanish in the limit $h \downarrow 0$, as $\tau < T(y) + h/r_1(j)$. Therefore, it only remains the constant of the probability $(h\mathbf{s}_+(j))/r_1(j)$.

The second type of paths happens when $\{X_t\}_{\geq 0}$ goes from $\mathcal{E}_+$ to $\mathcal{E}_0$ within the interval $[T(y), T(y) + (h/r_1(j))]$, and later on to the absorbing state $p+1$ without returning to $\mathcal{E}_+$. Here, $\{X_t\}_{\geq 0}$ jumps to $\mathcal{E}_0$ with probability $h\,\mathbf{S}_{+0}(j)/r_1(j)$, where $\mathbf{S}_{+0}(j)$ denotes the $j$-th row of $\mathbf{S}_{+0}$. Then, $\{X_t\}_{\geq 0}$ stays in $\mathcal{E}_0$ with transition probability matrix $\exp\{\mathbf{D}_0^{-1}\mathbf{S}_{00}\}$, and consequently goes to the absorbing state $p+1$ with a transition rate given by $\mathbf{s}_0$.

Putting all together, we conclude that for $h > 0$, which is small enough, we have

$$\begin{aligned} &\mathbb{E}\left(\exp\left\{-s\int_{T(y)}^{\tau} r_2\left(X_u\right) du\right\}, \mathcal{Y}_1 \leq y+h \,\middle|\, \mathcal{Y}_1 > y, X_{T(y)} = i\right) \\ =\; & h\,\mathbf{R}_+^{-1}\,\mathbf{s}_+ + h\,\mathbf{R}_+^{-1}\,\mathbf{S}_{+0}\left(s\mathbf{D}_0 - \mathbf{Q}_{00}\right)^{-1}\mathbf{s}_0 + o(h). \end{aligned}$$

Now, by taking the limit of the last expression it yields to

$$\mathbb{E}\left(e^{-s\mathcal{Y}_2}, \mathcal{Y}_1 \in dy \,\middle|\, X_0 = i\right) = \mathbf{e}_i^{\top}\exp\left\{\mathcal{W}(s)y\right\}\eta(s)dy, \tag{6.28}$$

for $y > 0$ and $i \in \mathcal{E}_+$.

Finally, by considering all the possible initial probabilities and using Theorem 6.11 and Equation (6.25), it results

$$\mathbb{E}\left(e^{-s\mathcal{Y}_2}, \mathcal{Y}_1 \in dy\right) = \left(\mathbf{s}_+ + \mathbf{s}_0\left(s\mathbf{D}_0 - \mathbf{S}_{00}\right)^{-1}\mathbf{S}_{0+}\right)\exp\left\{\mathcal{W}(s)y\right\}\eta(s)dy$$

for $y > 0$. □

As you may see, the matrix $\mathcal{W}(s)$ plays an important role in the semi-explicit formula for the density function and in the coming analysis we will need some properties of it and those are cited in the next Lemmas.

**Lemma 6.13** *For every $s \in \mathbb{C}$ with $\mathcal{R}e(s) \geq 0$, the spectrum of $\mathcal{W}(s)$ is contained in the open left-half plane.*

For the proof see [BOT05, Lemma 3].

**Lemma 6.14** *For every $s \in \mathbb{C}$ with $\mathcal{R}e(s) \geq 0$, we have $\chi(\mathcal{W}(s)) < 0$.*

The proof is given in [BO13, Lemma 3].

## 6.5 Distributions of concomitants of bivariate phase-type distributions

Let $(X, Y) \sim$ MPH* $(\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R})$ and let $h(x, y)$ denote its joint bivariate density function. Consider

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_{n+1}, Y_{n+1})$$

be a sample of i.i.d. pairs with common density $h(x, y)$.

Based on the ordering of the $Y$-variates, the purpose is to calculate the density of the $r$-th concomitant $X_{[r:n+1]}$.

First, we recall that

$$\mathbb{E}\left(e^{-sY}, X \in dx\right) := \mathbb{E}\left(e^{-sY}\mathbf{1}_{\{X \in dx\}}\right)$$

and thus we notice that

$$\mathbb{E}\left(e^{-sY}, X \in dx\right) = \int_0^\infty e^{-sy} h(y|X \in dx) dy f_X(x) dx, \tag{6.29}$$

where $f_X(x)$ denotes the density function of $X$.

Now, we recall Equation (6.26) and then from Equation (6.29) we obtain

$$\int_0^\infty e^{-ys} h(y|X \in dx) dy = \frac{\alpha(s)\, e^{\mathcal{W}(s)x} \eta(s)}{f_X(x)}, \tag{6.30}$$

for $x > 0$. Notice that the expression in the right side is analytic for all $\mathcal{R}e(s) \geq 0$ (we refer to the proof of Theorem 6.12 which is based on Theorem 6.5 and Lemma 6.6).

The general density function of the $r$-th concomitant (see Equation (6.5)) can be written as follows

$$
\begin{aligned}
f_{X_{[r:n+1]}}(x) &= r\binom{n+1}{r}\int_{-\infty}^{\infty}\left(1-\overline{F}_Y(y)\right)^{r-1}\left(\overline{F}_Y(y)\right)^{n+1-r}h(x,y)dy \\
&= r\binom{n+1}{r}\sum_{j=0}^{r-1}\binom{r-1}{j}(-1)^j\int_{-\infty}^{\infty}\left(\overline{F}_Y(y)\right)^{n+1-r+j}h(x,y)dy,
\end{aligned} \tag{6.31}
$$

where $\overline{F}_Y(y) = 1 - F_Y(y)$, and consequently we have

$$
f_{X_{[r:n+1]}}(x) = r\binom{n+1}{r}\sum_{j=0}^{r-1}\binom{r-1}{j}(-1)^j f_X(x)\int_{-\infty}^{\infty}\left(\overline{F}_Y(y)\right)^{u+j}h(y|X\in dx)dy, \tag{6.32}
$$

where $u = n+1-r$. In order to calculate the density function of the $r$-th concomitant, we need to calculate the integral

$$
\int_0^{\infty}\left(\overline{F}_Y(y)\right)^{m}h(y|X\in dx)dy. \tag{6.33}
$$

where $m = n+1-r+j$.

Let $(\boldsymbol{\beta}, \mathbf{T})$ be a PH-representation for the distribution of $Y$. Then, we get the following expression

$$
\left(\overline{F}_Y(y)\right)^m = \left(\boldsymbol{\beta}e^{\mathbf{T}y}\mathbf{e}\right)^m = \boldsymbol{\beta}^{\otimes m}e^{y\mathbf{T}^{\oplus m}}\mathbf{e}^{\otimes m} = \boldsymbol{\beta}^{\otimes m}(e^y)^{\mathbf{T}^{\oplus m}}\mathbf{e}^{\otimes m}.
$$

By using the last expression in Equation (6.33), we get

$$
\begin{aligned}
&\int_0^{\infty}\left(\overline{F}_Y(y)\right)^{m}h(y|X\in dx)dy \\
=\;& \int_0^{\infty}\boldsymbol{\beta}^{\otimes m}(e^y)^{\mathbf{T}^{\oplus m}}\mathbf{e}^{\otimes m}h(y|X\in dx)dy \\
=\;& \boldsymbol{\beta}^{\otimes m}\int_0^{\infty}(e^y)^{\mathbf{T}^{\oplus m}}h(y|X\in dx)dy\,\mathbf{e}^{\otimes m} \\
=\;& \boldsymbol{\beta}^{\otimes m}\int_0^{\infty}(e^{-y})^{-\mathbf{T}^{\oplus m}}h(y|X\in dx)dy\,\mathbf{e}^{\otimes m} \\
=\;& \boldsymbol{\beta}^{\otimes m}\int_0^{\infty}\frac{1}{2\pi i}\oint_{\gamma}e^{-yz}\left(z\mathbf{I}+\mathbf{T}^{\oplus m}\right)^{-1}dz h(y|X\in dx)dy\,\mathbf{e}^{\otimes m} \quad \mathcal{R}e(z)\geq 0,
\end{aligned}
$$

$$= \frac{\boldsymbol{\beta}^{\otimes m}}{2\pi i} \oint_{\gamma} \int_{0}^{\infty} e^{-zy} h(y|X \in dx)dy \left(z\mathbf{I} + \mathbf{T}^{\oplus m}\right)^{-1} dz\, \mathbf{e}^{\otimes m}. \tag{6.34}$$

Now, by substituting the expression in Equation (6.30) in to Equation (6.34), we have

$$\begin{aligned} &\int_{0}^{\infty} \left(\overline{F}_Y(y)\right)^m h(y|X \in dx)dy \\ = \;& \frac{\boldsymbol{\beta}^{\otimes m}}{f_X(x)} \frac{1}{2\pi i} \oint_{\gamma} \left(\alpha\left(z\right) e^{\mathcal{W}(z)x} \eta(z)\right) \left(z\mathbf{I} + \mathbf{T}^{\oplus m}\right)^{-1} dz\, \mathbf{e}^{\otimes m}. \end{aligned} \tag{6.35}$$

Here, notice that the spectrum of the matrix $-\mathbf{T}^{\oplus m}$ is contained in the open right-half plane.

In the next, we are going to explain how to calculate the integral

$$\frac{1}{2\pi i} \oint_{\gamma} \left(\alpha\left(z\right) e^{\mathcal{W}(z)x} \eta(z)\right) \left(z\mathbf{I} + \mathbf{T}^{\oplus m}\right)^{-1} dz. \tag{6.36}$$

Consider the Jordan canonical form of the matrix $-\mathbf{T}^{\oplus m}$ :

$$-\mathbf{T}^{\oplus m} = \mathbf{P}\mathbf{J}_{(m)}\mathbf{P}^{-1},$$

where

$$\mathbf{J}_{(m)} = \begin{pmatrix} \mathbf{J}_{(m,1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{(m,2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{J}_{(m,\kappa)} \end{pmatrix}$$

for some $\kappa \in \mathbb{N}$, while the matrix

$$\mathbf{J}_{(m,i)} = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{pmatrix}$$

represents the $i$-th Jordan block. Then we get

$$\left(z\mathbf{I} + \mathbf{T}^{\oplus m}\right)^{-1}$$

$$= \mathbf{P}\begin{pmatrix} \left(z\mathbf{I}_1 - \mathbf{J}_{(m,1)}\right)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \left(z\mathbf{I}_2 - \mathbf{J}_{(m,2)}\right)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \left(z\mathbf{I}_k - \mathbf{J}_{(m,\kappa)}\right)^{-1} \end{pmatrix}\mathbf{P}^{-1},$$

where

$$\left(z\mathbf{I}_i - \mathbf{J}_{(m,i)}\right)^{-1} = \begin{pmatrix} (z-\lambda_i)^{-1} & (z-\lambda_i)^{-2} & (z-\lambda_i)^{-3} & \cdots & (z-\lambda_i)^{-\epsilon_i} \\ 0 & (z-\lambda_i)^{-1} & (z-\lambda_i)^{-2} & \cdots & (z-\lambda_i)^{-(\epsilon_i-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & (z-\lambda_i)^{-1} \end{pmatrix} \tag{6.37}$$

for some $\epsilon_i \in \mathbb{N}$, and for every $i = 1, \ldots, \kappa$.

In this way, the integral in Equation (6.36) can be calculated as

$$\begin{aligned} &\frac{1}{2\pi i}\oint_\gamma \left(\alpha\left(z\right)e^{\mathcal{W}(z)x}\eta(z)\right)\left(z\mathbf{I} + \mathbf{T}^{\oplus m}\right)^{-1} dz \\ = \quad & \mathbf{P}\frac{1}{2\pi i}\oint_\gamma \left(\alpha\left(z\right)e^{\mathcal{W}(z)x}\eta(z)\right)\left(z\mathbf{I} - \mathbf{J}_{(m)}\right)^{-1} dz\mathbf{P}^{-1} \end{aligned}$$

and in particular

$$\begin{aligned} &\frac{1}{2\pi i}\oint_\gamma \left(\alpha\left(z\right)e^{\mathcal{W}(z)x}\eta(z)\right)\left(z\mathbf{I} - \mathbf{J}_{(m)}\right)^{-1} dz \\ = \quad & \bigoplus_{i=1}^{\kappa}\frac{1}{2\pi i}\oint_\gamma \left(\alpha\left(z\right)e^{\mathcal{W}(z)x}\eta(z)\right)\left(z\mathbf{I}_1 - \mathbf{J}_{(m,i)}\right)^{-1} dz, \end{aligned}$$

where $\bigoplus$ denotes the direct sum.

Lastly, in order to calculate

$$\tfrac{1}{2\pi i}\oint_\gamma \alpha\left(z\right)e^{\mathcal{W}(z)x}\eta(z)\left(z\mathbf{I}_i - \mathbf{J}_{(m,i)}\right)^{-1} dz$$

for every $i = 1, \ldots, \kappa$, let us denote $f(z,x) = \alpha\left(z\right)e^{\mathcal{W}(z)x}\eta(z)$. Then

$$\frac{1}{2\pi i}\oint_\gamma f(z,x)\left(z\mathbf{I}_i - \mathbf{J}_{(m,i)}\right)^{-1} dz$$

$$= \begin{pmatrix} f(\lambda_i, x) & \frac{\partial f(z,x)}{\partial z}\Big|_{z=\lambda_i} & \cdots & \frac{1}{\epsilon_i!} \frac{\partial^{\epsilon_i} f(z,x)}{\partial z^{\epsilon_i}}\Big|_{z=\lambda_i} \\ 0 & f(\lambda_i, x) & \cdots & \frac{1}{(\epsilon_i - 1)!} \frac{\partial^{(\epsilon_i - 1)} f(z,x)}{\partial z^{(\epsilon_i - 1)}}\Big|_{z=\lambda_i} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_i, x) \end{pmatrix}$$

$$=: f(\mathbf{J}_{(m,i)}, x).$$

Consequently, we can write

$$\begin{aligned} & \frac{1}{2\pi i} \oint_\gamma f(z,x) \left(z\mathbf{I} - \mathbf{J}_{(m)}\right)^{-1} dz \\ = & \begin{pmatrix} f(\mathbf{J}_{(m,1)}, x) & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & f(\mathbf{J}_{(m,\kappa)}, x) \end{pmatrix} \\ = & \bigoplus_{i=1}^{\kappa} f(\mathbf{J}_{(m,i)}, x) \end{aligned}$$

and we get

$$\tfrac{1}{2\pi i} \oint_\gamma f(z,x) \left(z\mathbf{I} + \mathbf{T}^{\oplus m}\right)^{-1} dz = \mathbf{P} \bigoplus_{i=1}^{\kappa} f(\mathbf{J}_{(m,i)}, x) \mathbf{P}^{-1}.$$

Finally, by combining Equations (6.35) and (6.32), we have shown a way to calculate the density function of the $r$-th concomitant.

In following part we are going to derive a closed form formula for the density function of the $r$-th concomitant which is going to help to prove that concomitants are phase-type distributed.

consider the following matrices

$$\alpha\big(\mathbf{J}_{(m,i)}\big) := \begin{pmatrix} \alpha(\lambda_i) & \frac{\partial\, \alpha(z)}{\partial z}\Big|_{z=\lambda_i} & \cdots & \frac{1}{\epsilon_i!} \frac{\partial^{\epsilon_i}\, \alpha(z)}{\partial z^{\epsilon_i}}\Big|_{z=\lambda_i} \\ \mathbf{0} & \alpha(\lambda_i) & \cdots & \frac{1}{(\epsilon_i - 1)!} \frac{\partial^{(\epsilon_i - 1)}\, \alpha(z)}{\partial z^{(\epsilon_i - 1)}}\Big|_{z=\lambda_i} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \alpha(\lambda_i) \end{pmatrix}$$

$$
e^{\mathcal{W}(\mathbf{J}_{(m,i)})x} := \begin{pmatrix} e^{\mathcal{W}(\lambda_i)x} & \left.\frac{\partial\, e^{\mathcal{W}(z)x}}{\partial z}\right|_{z=\lambda_i} & \cdots & \left.\frac{1}{\epsilon_i!}\frac{\partial^{\epsilon_i}\, e^{\mathcal{W}(z)x}}{\partial z^{\epsilon_i}}\right|_{z=\lambda_i} \\ \mathbf{0} & e^{\mathcal{W}(\lambda_i)x} & \cdots & \left.\frac{1}{(\epsilon_i-1)!}\frac{\partial^{(\epsilon_i-1)}\, e^{\mathcal{W}(z)x}}{\partial z^{(\epsilon_i-1)}}\right|_{z=\lambda_i} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & e^{\mathcal{W}(\lambda_i)x} \end{pmatrix}
$$

$$
\eta(\mathbf{J}_{(m,i)}) := \begin{pmatrix} \eta(\lambda_i) & \left.\frac{\partial\, \eta(z)}{\partial z}\right|_{z=\lambda_i} & \cdots & \left.\frac{1}{\epsilon_i!}\frac{\partial^{\epsilon_i}\, \eta(z)}{\partial z^{\epsilon_i}}\right|_{z=\lambda_i} \\ \mathbf{0} & \eta(\lambda_i) & \cdots & \left.\frac{1}{(\epsilon_i-1)!}\frac{\partial^{(\epsilon_i-1)}\, \eta(z)}{\partial z^{(\epsilon_i-1)}}\right|_{z=\lambda_i} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \eta(\lambda_i) \end{pmatrix}.
$$

Then, we have

$$
f(\mathbf{J}_{(m,i)}, x) = \alpha(\mathbf{J}_{(m,i)})e^{\mathcal{W}(\mathbf{J}_{(m,i)})x}\eta(\mathbf{J}_{(m,i)})
$$

for every $i = 1, \ldots, \kappa_m$, and consequently we have

$$
\frac{1}{2\pi i}\oint_{\gamma} f(z,x)\left(z\mathbf{I} - \mathbf{J}_{(m)}\right)^{-1} dz = \bigoplus_{i=1}^{\kappa}\left(\alpha(\mathbf{J}_{(m,i)})e^{\mathcal{W}(\mathbf{J}_{(m,i)})x}\eta(\mathbf{J}_{(m,i)})\right).
$$

Now, let us introduce the following matrices

$$
\mathbf{\Lambda}_{(m)} = \begin{pmatrix} \alpha(\mathbf{J}_{(m,1)}) & \mathbf{0} & \cdots & 0 \\ 0 & \alpha(\mathbf{J}_{(m,2)}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha(\mathbf{J}_{(m,k)}) \end{pmatrix},
$$

$$
\mathbf{W}(\mathbf{J}_{(m)}, x) = \begin{pmatrix} e^{\mathcal{W}(\mathbf{J}_{(m,1)})x} & 0 & \cdots & 0 \\ 0 & e^{\mathcal{W}(\mathbf{J}_{(m,2)})x} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\mathcal{W}(\mathbf{J}_{(m,\kappa)})x} \end{pmatrix},
$$

$$\mathbf{N}_{(m)} = \begin{pmatrix} \eta(\mathbf{J}_{(m,1)}) & 0 & \cdots & 0 \\ 0 & \eta(\mathbf{J}_{(m,2)}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \eta(\mathbf{J}_{(m,\kappa)}) \end{pmatrix}.$$

By using the latest notation for the matrices, we can write

$$\begin{aligned} & \frac{1}{2\pi i}\oint_{\gamma}\left(\alpha\left(z\right)e^{\mathcal{W}(z)x}\eta(z)\right)\left(z\mathbf{I}-\mathbf{J}_{(m)}\right)^{-1}dz \\ = \; & \mathbf{\Lambda}_{(m)}\mathbf{W}(\mathbf{J}_{(m)},x)\mathbf{N}_{(m)}. \end{aligned} \tag{6.38}$$

Then, we get

$$\frac{1}{2\pi i}\oint_{\gamma}\left(\alpha\left(z\right)e^{\mathcal{W}(z)x}\eta(z)\right)\left(z\mathbf{I}+\mathbf{T}^{\oplus m}\right)^{-1}dz = \mathbf{P}\mathbf{\Lambda}_{(m)}\mathbf{W}(\mathbf{J}_{(m)},x)\mathbf{N}_{(m)}\mathbf{P}^{-1}.$$

Therefore, the integral in Equation (6.35) is given by

$$\int_0^{\infty}\left(\overline{F}_Y(y)\right)^m h(y|X\in dx)dy = \frac{\boldsymbol{\beta}^{\otimes m}}{f_X(x)}\mathbf{P}\mathbf{\Lambda}_{(m)}\mathbf{W}(\mathbf{J}_{(m)},x)\mathbf{N}_{(m)}\mathbf{P}^{-1}\mathbf{e}^{\otimes m}. \tag{6.39}$$

Now, by combining Equation (6.32) and Equation (6.39), we get the following expression for the density of the $r$-th concomitant.

$$\begin{aligned} & f_{X_{[r:n+1]}}(x) = \\ = \; & r\tbinom{n+1}{r}\sum_{j=0}^{r-1}\tbinom{r-1}{j}(-1)^j f_X(x)\int_0^{\infty}\left(\overline{F}_Y(y)\right)^{u+j}h(y|X\in dx)dy \\ = \; & r\tbinom{n+1}{r}\sum_{j=0}^{r-1}\tbinom{r-1}{j}(-1)^j\boldsymbol{\beta}^{\otimes u+j}\mathbf{P}\mathbf{\Lambda}_{(u+j)}\mathbf{W}(\mathbf{J}_{(u+j)},x)\mathbf{N}_{(u+j)}\mathbf{P}^{-1}\mathbf{e}^{\otimes u+j}, \end{aligned} \tag{6.40}$$

where $u = n-(r-1)$.

Let us define the vectors $\boldsymbol{\gamma}_r = \left(\boldsymbol{\gamma}_0,\ldots,\boldsymbol{\gamma}_{r-1}\right)$ and $\mathbf{g}_r = \left(\mathbf{g}_0,\ldots,\mathbf{g}_{r-1}\right)$, where

$$\boldsymbol{\gamma}_j = r\tbinom{n+1}{r}\tbinom{r-1}{j}(-1)^j\boldsymbol{\beta}^{\otimes u+j}\mathbf{P}\mathbf{\Lambda}_{(u+j)},$$

$$\mathbf{g}_j = \mathbf{N}_{(u+j)}\mathbf{P}^{-1}\mathbf{e}^{\otimes u+j},$$

for every $j = 0, \ldots, r-1$.

Also, we define the matrix

$$\mathbf{G}_r(x) = \bigoplus_{j=0}^{r-1} \mathbf{W}\left(\mathbf{J}_{(u+j)}, x\right).$$

Finally, the density for the $r$-th concomitant can be written as

$$f_{X_{[r:n+1]}}(x) = \boldsymbol{\gamma}_r \mathbf{G}_r(x)\mathbf{g}_r. \tag{6.41}$$

Now, we introduce the following two lemmas to prove that concomitants are phase-type distributed according to Theorem 3.4.

**Lemma 6.15** *The density function $f_{X_{[r:n+1]}}(x)$ is positive on the positive reals.*

**Proof.** From formula (6.31) we can see that if $h(x, y) = 0$, then $f_{X_{[r:n+1]}}(x) = 0$. Assume that there exists $x^* > 0$ such that $h(x^*, y) = 0$. Then, $\int_0^\infty h(x^*, y)dy = f_X(x^*) = 0$, which is a contradiction since $X$ is phase-type distributed (recall Proposition 6.10 and Theorem 3.4). □

**Lemma 6.16** *The Laplace transform of $f_{X_{[r:n+1]}}(x)$, denoted by $\mathcal{L}_{f_{X_{[r:n+1]}}}(s)$, is a rational function and it has a unique pole of maximal real part.*

**Proof.**

$$\begin{aligned}
\mathcal{L}_{f_{X_{[r:n+1]}}}(s) &= \int_0^\infty e^{-sx} f_{X_{[r:n+1]}}(x)dx \\
&= \boldsymbol{\gamma}_r \int_0^\infty e^{-sx}\mathbf{G}_r(x)dx\mathbf{g}_r \\
&= \boldsymbol{\gamma}_r \bigoplus_{j=0}^{r-1}\left(\int_0^\infty e^{-sx}\mathbf{W}\left(\mathbf{J}_{(u+j)}, x\right)dx\right)\mathbf{g}_r, \quad u = n-(r-1), \\
&= \boldsymbol{\gamma}_r \bigoplus_{j=0}^{r-1}\left(\bigoplus_{i=1}^{\kappa_j}\left(\int_0^\infty e^{-sx} e^{\mathcal{W}\left(\mathbf{J}_{(u+j,i)}\right)x}dx\right)\right)\mathbf{g}_r.
\end{aligned}$$

Notice that

$$e^{\mathcal{W}\left(\mathbf{J}_{(u+j,i)}\right)x} =$$

$$\begin{pmatrix} e^{\mathcal{W}(\lambda_i)x} & \left.\frac{\partial\, e^{\mathcal{W}(z)x}}{\partial z}\right|_{z=\lambda_i} & \cdots & \left.\frac{1}{\epsilon_i!}\frac{\partial^{\epsilon_i}\, e^{\mathcal{W}(z)x}}{\partial z^{\epsilon_i}}\right|_{z=\lambda_i} \\ \mathbf{0} & e^{\mathcal{W}(\lambda_i)x} & \cdots & \left.\frac{1}{(\epsilon_i-1)!}\frac{\partial^{(\epsilon_i-1)}\, e^{\mathcal{W}(z)x}}{\partial z^{(\epsilon_i-1)}}\right|_{z=\lambda_i} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & e^{\mathcal{W}(\lambda_i)x} \end{pmatrix}.$$

Then, we need to calculate

$$\int_0^\infty e^{-sx}\left(\frac{1}{k!}\left.\frac{\partial^k\, e^{\mathcal{W}(z)x}}{\partial z^k}\right|_{z=\lambda_i}\right)dx, \tag{6.42}$$

for all $k = 1, \ldots, \epsilon_i$.

Now, by using the *Faà di Bruno's formula* (see Equation (A.7)), we can write the $k$-th derivative as follows

$$\left.\frac{\partial^k\, e^{\mathcal{W}(z)x}}{\partial z^k}\right|_{z=\lambda_i} = \sum_{\ell=1}^{k}\mathcal{G}_{k,\ell}\left(\mathcal{W}(\lambda_i)\right)x^\ell e^{\mathcal{W}(\lambda_i)x}, \quad k \geq 1, \tag{6.43}$$

where the function $\mathcal{G}_{k,\ell}\left(\mathcal{W}(\lambda_i)\right)$ is given by the derivatives of the function $\mathcal{W}(\lambda_i)$ with respect to $\lambda_i$ and which derivatives have coefficient $x^\ell$.

By substituting Equation (6.43) into Equation (6.42), we get

$$\begin{aligned}\frac{1}{k!}\int_0^\infty e^{-sx}\left(\left.\frac{\partial^k\, e^{\mathcal{W}(z)x}}{\partial z^k}\right|_{z=\lambda_i}\right)dx &= \frac{1}{k!}\sum_{\ell=1}^{k}\mathcal{G}_{k,\ell}\left(\mathcal{W}(\lambda_i)\right)\int_0^\infty x^\ell e^{-sx}e^{\mathcal{W}(\lambda_i)x}dx \\ &= \frac{1}{k!}\sum_{\ell=1}^{k}\mathcal{G}_{k,\ell}\left(\mathcal{W}(\lambda_i)\right)\int_0^\infty x^\ell e^{-(s\mathbf{I}-\mathcal{W}(\lambda_i))x}dx \\ &= \frac{1}{k!}\sum_{\ell=1}^{k}\ell!\,\mathcal{G}_{k,\ell}\left(\mathcal{W}(\lambda_i)\right)\left(s\mathbf{I}-\mathcal{W}(\lambda_i)\right)^{-\ell-1}, \end{aligned} \tag{6.44}$$

for every $k = 1, \ldots, \epsilon_i$.

Going back to the integral

$$\int_0^\infty e^{-sx}e^{\mathcal{W}\left(\mathbf{J}_{(u+j,i)}\right)x}dx,$$

this integral is calculated as follows

$$\begin{pmatrix} (s\mathbf{I}-\mathcal{W}(\lambda_i))^{-1} & \left(\frac{\partial\,\mathcal{W}(z)}{\partial z}\Big|_{z=\lambda_i}\right)(s\mathbf{I}-\mathcal{W}(\lambda_i))^{-2} & \cdots & \frac{1}{\epsilon_i!}\sum_{\ell=1}^{\epsilon_i}\ell!\,\mathcal{G}_{\epsilon_i,\ell}\left(\mathcal{W}(\lambda_i)\right)(s\mathbf{I}-\mathcal{W}(\lambda_i))^{-\ell-1} \\ \mathbf{0} & (s\mathbf{I}-\mathcal{W}(\lambda_i))^{-1} & \cdots & \frac{1}{(\epsilon_i-1)!}\sum_{\ell=1}^{\epsilon_i-1}\ell!\,\mathcal{G}_{\epsilon_i-1,\ell}\left(\mathcal{W}(\lambda_i)\right)(s\mathbf{I}-\mathcal{W}(\lambda_i))^{-\ell-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (s\mathbf{I}-\mathcal{W}(\lambda_i))^{-1} \end{pmatrix}. \tag{6.45}$$

Let us denote

$$\int_0^\infty e^{-sx}e^{\mathcal{W}\left(\mathbf{J}_{(u+j,i)}\right)x}dx =: \left(s\mathbf{I}-\mathcal{W}\left(\mathbf{J}_{(u+j,i)}\right)\right)^{-1}.$$

Now, consider the product

$$\alpha\left(\mathbf{J}_{(u+j,i)}\right)\left(s\mathbf{I}-\mathcal{W}\left(\mathbf{J}_{(u+j,i)}\right)\right)^{-1}\eta\left(\mathbf{J}_{(u+j,i)}\right),$$

then the direct sum

$$\bigoplus_{i=1}^{\kappa_j}\alpha\left(\mathbf{J}_{(u+j,i)}\right)\left(s\mathbf{I}-\mathcal{W}\left(\mathbf{J}_{(u+j,i)}\right)\right)^{-1}\eta\left(\mathbf{J}_{(u+j,i)}\right),$$

which consequently yields us to

$$\mathcal{L}_{f_{X_{[r:n+1]}}}(s) = \boldsymbol{\gamma}_r\bigoplus_{j=0}^{r-1}\left(\bigoplus_{i=1}^{\kappa_j}\left(s\mathbf{I}-\mathcal{W}\left(\mathbf{J}_{(u+j,i)}\right)\right)^{-1}\right)\mathbf{g}_r. \tag{6.46}$$

From Equation (6.44) we can see that the Laplace transform $\mathcal{L}_{f_{X_{[r:n+1]}}}(s)$ consists on a partial fraction expansion of terms with the form

$$\frac{1}{k!}\sum_{\ell=1}^{k}\ell!\,\mathcal{G}_{k,\ell}\left(\mathcal{W}(\lambda_i)\right)(s\mathbf{I}-\mathcal{W}(\lambda_i))^{-\ell-1}$$

which are in function of the eigenvalues of the matrices $\mathbf{T}^{\oplus(u+j)}$, where $j=0,1,\ldots,r-1$. Hence, $\mathcal{L}_{f_{X_{[r:n+1]}}}(s)$ is a rational function.

In the next part, we analyse the poles of $\mathcal{L}_{f_{X_{[r:n+1]}}}(s)$.

From the compact form of the Laplace transform $\mathcal{L}_{f_{X_{[r:n+1]}}}(s)$ given in Equation (6.46) and from the matrix in (6.45), the first observation is that all possible poles of $\mathcal{L}_{f_{X_{[r:n+1]}}}(s)$ are given by the poles of all terms of the form

$$(s\mathbf{I} - \mathcal{W}(\lambda_i))^{-1}$$

and those poles are the eigenvalues of the corresponding matrix $\mathcal{W}(\lambda_i)$ (see Equation (6.37)).

Let $\lambda_{\max}$ denote the eigenvalue of the sub-intensity matrix $\mathbf{T}$ with the maximum real part (which is negative). Then $m\lambda_{\max}$ is the eigenvalue of $\mathbf{T}^{\oplus m}$ with the maximum real part (which is negative). Consequently, $-m\lambda_{\max}$ is the eigenvalue of $-\mathbf{T}^{\oplus m}$ with the smallest real part (which is positive).

Consider the matrix $-\mathbf{T}^{\oplus(n+1-r)}$ and its eigenvalue with the smallest real part $-(n+1-r)\lambda_{\max}$, which we are going to denote by $s_r^*$, ( this is $s_r^* = -(n+1-r)\lambda_{\max}$). Next, we are going to verify that the maximum real part eigenvalue of the matrix $\mathcal{W}(s_r^*)$ is exactly the pole of $\mathcal{L}_{f_{X_{[r:n+1]}}}(s)$ with the maximum real part.

Let us consider the second smallest eigenvalue of $-\mathbf{T}^{\oplus(n+1-r)}$. We get this eigenvalue by considering the second largest eigenvalue of $\mathbf{T}$. Let us denote by $\lambda^{(2)}$ to the second largest eigenvalue of $\mathbf{T}$. Thus

$$\lambda^{(2)} < \lambda_{\max}$$

and consequently

$$-(n+1-r)\lambda_{\max} < -(n+r)\lambda_{\max} - \lambda^{(2)}.$$

We denote $-(n+r)\lambda_{\max} - \lambda^{(2)} =: \tilde{s}_r$. Hence, $s_r^* < \tilde{s}_r$.

We are going to compare the eigenvalues of the matrices $\mathcal{W}(s_r^*)$ and $\mathcal{W}(\tilde{s}_r)$.

We recall that for a matrix $\mathbf{A}$, $\chi(\mathbf{A})$ denotes the maximal real eigenvalue of $\mathbf{A}$.

$\mathcal{W}(s)$ is the generator of the uniformly continuous semi-group given by $\Delta^y(s) = e^{\mathcal{W}(s)y}$ where $y \geq 0$ and $\mathcal{R}e(s) \geq 0$ (see Lemma 1 in [BOT05]). Also, $\chi(\Delta^y(s))$ is a decreasing function with respect to $s$ (see [LRS+12, p.199] and [Gau97, Section 2.2.3]). Thus,

$$\chi(\Delta^y(s_r^*)) > \chi(\Delta^y(\tilde{s}_r)). \tag{6.47}$$

Now, by the spectral mapping theorem we have that

$$\chi\left(e^{\mathcal{W}(s_r^*)y}\right) = e^{y\varsigma^*},$$

where $\varsigma^* \in \sigma\left(\mathcal{W}(s_r^*)\right)$, and $\sigma\left(\mathcal{W}(s_r^*)\right)$ denotes the spectrum of $\mathcal{W}(s_r^*)$. As well as we have

$$\chi\left(e^{\mathcal{W}(\tilde{s}_r)y}\right) = e^{y\tilde{\varepsilon}},$$

where $\tilde{\varepsilon} \in \sigma\left(\mathcal{W}(s_2)\right)$. Then, by relation (6.47), it follows that

$$\varsigma^* > \tilde{\varepsilon}.$$

Therefore, we conclude that $\varsigma^*$ is the pole with the maximal real part and $\varsigma^*$ is negative (see Lemma 6.7). $\square$

### 6.5.1 Example: concomitants from a sample of bivariate Erlang distributed random vectors.

We recall Example 6.3.1. Here, we consider

$$X = \frac{Z}{\lambda_1}, \quad Y = \frac{Z+W}{\lambda_2},$$

where the variables $Z$ and $W$ are independent and exponentially distributed with intensity 1. Then, $(X,Y) \sim \text{MPH}^*\left(\boldsymbol{\alpha}, \mathbf{S}; \mathbf{R}\right)$, where

$$\boldsymbol{\alpha} = (1,0), \qquad \mathbf{S} = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad \mathbf{R} = \begin{pmatrix} \frac{1}{\lambda_1} & \frac{1}{\lambda_2} \\ 0 & \frac{1}{\lambda_2} \end{pmatrix},$$

and we have that $X \sim \text{Erlang}\,(1,\lambda_1)$ and $Y \sim \text{Erlang}\,(2,\lambda_2) \sim \text{PH}\,(\boldsymbol{\beta}, \mathbf{T})$, where

$$\boldsymbol{\beta} = (1,0) \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} -\lambda_2 & \lambda_2 \\ 0 & -\lambda_2 \end{pmatrix},$$

see Example 3.2.

**Case:** $n = 2$.
Consider a sample of two elements of this bivariate distribution:

$$\left(X_1, Y_1\right), \left(X_2, Y_2\right)$$

and consider the ordering with respect to the $Y$-variate:

$$\left(X_{[1:2]}, Y_{(1:2)}\right), \left(X_{[2:2]}, Y_{(2:2)}\right).$$

Our purpose is to determine the density function of the first and second concomitant.

We start with the concomitant of the minimum, this is the variable $X_{[1:2]}$. According to the Equation (6.40), the density of $X_{[1:2]}$ can be calculated as follows

$$\begin{aligned} f_{X_{[1:2]}}(x) &= 2\boldsymbol{\beta}\frac{1}{2\pi i}\oint_{\gamma}\alpha(z)e^{\mathcal{W}(z)x}\eta(z)\left(z\mathbf{I}+\mathbf{T}\right)^{-1}dz\mathbf{e} \\ &= 2\boldsymbol{\beta}\mathbf{P}\frac{1}{2\pi i}\oint_{\gamma}\alpha(z)e^{\mathcal{W}(z)x}\eta(z)\left(z\mathbf{I}-\mathbf{J}_{1}\right)^{-1}dz\mathbf{P}^{-1}\mathbf{e} \end{aligned}$$

where

$$\mathbf{J}_{1}=\begin{pmatrix}\lambda_2 & 1\\ 0 & \lambda_2\end{pmatrix} \quad \text{and} \quad \mathbf{P}=\begin{pmatrix}-\lambda_2 & 0\\ 0 & 1\end{pmatrix}.$$

Consequently,

$$f_{X_{[1:2]}}(x) = 2\boldsymbol{\beta}\mathbf{P}\begin{pmatrix}\alpha(\lambda_2)e^{\mathcal{W}(\lambda_2)x}\eta(\lambda_2) & \frac{\partial\left(\alpha(\lambda_2)e^{\mathcal{W}(\lambda_2)x}\eta(\lambda_2)\right)}{\partial\lambda_2}\\ 0 & \alpha(\lambda_2)e^{\mathcal{W}(\lambda_2)x}\eta(\lambda_2)\end{pmatrix}\mathbf{P}^{-1}\mathbf{e}.$$

Note that for this particular representation $(\boldsymbol{\alpha},\mathbf{S};\mathbf{R})$ we have

$$\begin{aligned} &\alpha_{+}=1, \quad \alpha_0=0,\\ &\mathbf{S}_{00}=-1, \quad \mathbf{S}_{0+}=0, \quad \mathbf{S}_{+0}=1, \quad \mathbf{S}_{++}=-1,\\ &\mathbf{R}_{+}=\frac{1}{\lambda_1}, \quad \mathbf{D}_{+}=\frac{1}{\lambda_2}, \quad \mathbf{D}_0=\frac{1}{\lambda_2},\\ &\eta_0=1, \quad \eta_{+}=0. \end{aligned}$$

Then,

$$\alpha(z)=1, \qquad \mathcal{W}(z)=-\lambda_1\left(\frac{\lambda_2+z}{\lambda_2}\right) \quad \text{and} \quad \eta(z)=\frac{\lambda_1\lambda_2}{z+\lambda_2}.$$

Thus,

$$\alpha(z)e^{\mathcal{W}(z)x}\eta(z)=\frac{\lambda_1\lambda_2}{z+\lambda_2}e^{-\lambda_1\left(\frac{\lambda_2+z}{\lambda_2}\right)x}=:g(z),$$

and then

$$\frac{\partial g(z)}{\partial z}=\left(\frac{-\lambda_1\lambda_2-\lambda_1^2x(z+\lambda_2)}{(z+\lambda_2)^2}\right)e^{-\lambda_1\left(\frac{\lambda_2+z}{\lambda_2}\right)x}.$$

Consequently, we have

$$\alpha(\lambda_2)e^{\mathcal{W}(\lambda_2)x}\eta(\lambda_2)=\frac{\lambda_1}{2}e^{-2\lambda_1x}$$

and

$$\frac{\partial\, g(\lambda_2)}{\partial \lambda_2} = -\frac{\lambda_1^2 x}{2\lambda_2} e^{-2\lambda_1 x} - \frac{\lambda_1}{4\lambda_2} e^{-2\lambda_1 x}.$$

Finally, we obtain

$$\begin{aligned}
f_{X_{[1:2]}}(x) &= 2\boldsymbol{\beta}\mathbf{P} \begin{pmatrix} \frac{\lambda_1}{2} e^{-2\lambda_1 x} & -\frac{\lambda_1^2 x}{2\lambda_2} e^{-2\lambda_1 x} - \frac{\lambda_1}{4\lambda_2} e^{-2\lambda_1 x} \\ 0 & \frac{\lambda_1}{2} e^{-2\lambda_1 x} \end{pmatrix} \mathbf{P}^{-1}\mathbf{e} \\
&= 2\boldsymbol{\beta} \begin{pmatrix} \frac{\lambda_1}{2} e^{-2\lambda_1 x} & \frac{\lambda_1^2 x}{2} e^{-2\lambda_1 x} + \frac{\lambda_1}{4} e^{-2\lambda_1 x} \\ 0 & \frac{\lambda_1}{2} e^{-2\lambda_1 x} \end{pmatrix} \mathbf{e} \\
&= 2\left(\frac{3\lambda_1}{4} + \frac{\lambda_1^2 x}{2}\right) e^{-2\lambda_1 x} \\
&= \lambda_1 \left(\frac{3}{2} + \lambda_1 x\right) e^{-2\lambda_1 x}.
\end{aligned}$$

A PH-representation for the distribution of $X_{[1:2]}$ is given by $\left(\boldsymbol{\alpha}_{[1:2]}, \mathbf{S}_{[1:2]}\right)$, where the initial vector is given by

$$\boldsymbol{\alpha}_{[1:2]} = (3/4, 1/4, 0)$$

and the sub-intensity matrix

$$\mathbf{S}_{[1:2]} = \begin{pmatrix} -2\lambda_1 & 0 & 0 \\ 0 & -2\lambda_1 & 2\lambda_1 \\ 0 & 0 & -2\lambda_1 \end{pmatrix}.$$

Therefore, we can see $X_{[1:2]}$ as a mixture of an Exponential distributed random variable with intensity $2\lambda_1$ and an Erlang distributed random variable with parameters $(2, 2\lambda_1)$.

Now, we calculate the density of the concomitant of the maximum $X_{[2:2]}$. This is is given by

$$\begin{aligned}
f_{X_{[2:2]}}(x) &= 2 f_X(x) - f_{X_{[1:2]}}(x) \\
&= 2\lambda_1 e^{-\lambda_1 x} - \lambda_1 \left(\frac{3}{2} + \lambda_1 x\right) e^{-2\lambda_1 x},
\end{aligned}$$

see Equation (6.40).

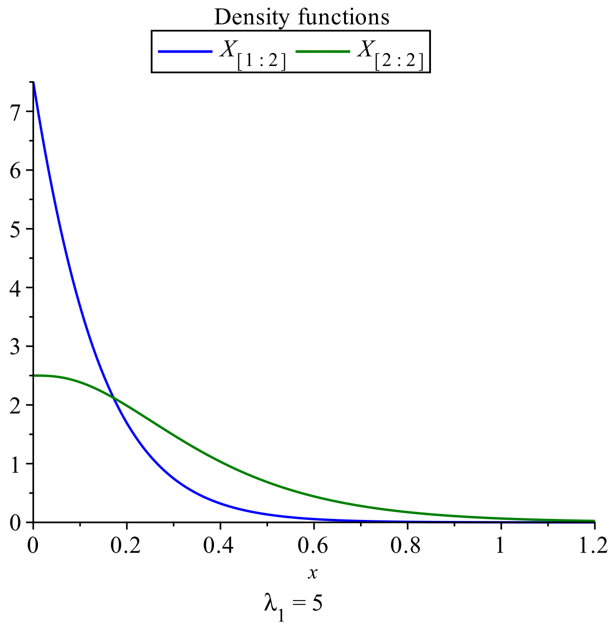
A PH-representation for the distribution of $X_{[2:2]}$ is given by $\left(\boldsymbol{\alpha}_{[2:2]}, \mathbf{S}_{[2:2]}\right)$, where the initial vector is

$$\boldsymbol{\alpha}_{[2:2]} = (1/2, 0, 0, 1/4, 0, 1/4).$$

and the sub-intensity matrix is

$$\mathbf{S}_{[2:2]} = \begin{pmatrix} -2\lambda_1 & 2\lambda_1 & 0 & 0 & 0 & 0 \\ 0 & -2\lambda_1 & 2\lambda_1 & 0 & 0 & 0 \\ 0 & 0 & -\lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2\lambda_1 & 2\lambda_1 & 0 \\ 0 & 0 & 0 & 0 & -2\lambda_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2\lambda_1 \end{pmatrix}.$$

Thus, the distribution of $X_{[2:2]}$ can be seen as the mixture of a Coxian, Erlang and an Exponential distributed random variables.




**Case** $n = 3$.
Now we consider a sample of three elements of the same bivariate distribution:

$$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$$

and consider again the ordering with respect to the $Y$-variate:

$$\left(X_{[1:3]}, Y_{(1:3)}\right), \left(X_{[2:3]}, Y_{(2:3)}\right) \left(X_{[3:3]}, Y_{(3:3)}\right).$$

Then, by using Equation (6.40), the density functions of the concomitants are

$$f_{X_{[1:3]}}(x) = \lambda_1 \left(\frac{17}{9} + \frac{8}{3}\lambda_1 x + \lambda_1^2 x^2\right) e^{-3\lambda_1 x},$$

$$
\begin{aligned}
f_{X_{[2:3]}}(x) &= 3\lambda_1\left(\frac{3}{2}+\lambda_1 x\right)e^{-2\lambda_1 x}-2\lambda_1\left(\frac{17}{9}+\frac{8}{3}\lambda_1 x+\lambda_1^2x^2\right)e^{-3\lambda_1 x},\\
f_{X_{[3:3]}}(x) &= 3\lambda_1 e^{-\lambda_1 x}-3\lambda_1\left(\frac{3}{2}+\lambda_1 x\right)e^{-2\lambda_1 x}+\lambda_1\left(\frac{17}{9}+\frac{8}{3}\lambda_1 x+\lambda_1^2x^2\right)e^{-3\lambda_1 x}.
\end{aligned}
$$

A PH-representation for the distribution of $X_{[1:3]}$ is given by $\left(\boldsymbol{\alpha}_{[1:3]}, \mathbf{S}_{1:3}\right)$, where the vector of initial probabilities is

$$\boldsymbol{\alpha}_{[1:3]} = (17/27, 8/27, 0, 2/27, 0, 0)$$

the sub-intensity matrix is

$$
\mathbf{S}_{[1:3]} = \begin{pmatrix}
-3\lambda_1 & 0 & 0 & 0 & 0 & 0\\
0 & -3\lambda_1 & 3\lambda_1 & 0 & 0 & 0\\
0 & 0 & -3\lambda_1 & 0 & 0 & 0\\
0 & 0 & 0 & -3\lambda_1 & 3\lambda_1 & 0\\
0 & 0 & 0 & 0 & -3\lambda_1 & 3\lambda_1\\
0 & 0 & 0 & 0 & 0 & -3\lambda_1
\end{pmatrix}.
$$

Now, a PH-representation for the distribution of $X_{[2:3]}$ is denoted by $\left(\boldsymbol{\alpha}_{[2:3]}, \mathbf{S}_{[2:3]}\right)$, where the vector of initial probabilities is

$$\boldsymbol{\alpha}_{[2:3]} = \left(\boldsymbol{\beta}_{(1)}, \boldsymbol{\beta}_{(2)}, \boldsymbol{\beta}_{(3)}, \boldsymbol{\beta}_{(4)}, \boldsymbol{\beta}_{(5)}, \boldsymbol{\beta}_{(6)}\right),$$

where

$$
\begin{aligned}
\boldsymbol{\beta}_{(1)} &= \frac{13}{54}, \qquad \boldsymbol{\beta}_{(2)} = \left(\frac{13}{54}, 0\right), \qquad \boldsymbol{\beta}_{(3)} = \left(\frac{2}{27}, 0, 0\right),\\
\boldsymbol{\beta}_{(4)} &= \left(\frac{1}{4}, 0, 0\right), \qquad \boldsymbol{\beta}_{(5)} = \left(\frac{1}{9}, 0, 0, 0\right), \qquad \boldsymbol{\beta}_{(6)} = \left(\frac{1}{12}, 0, 0, 0\right).
\end{aligned}
$$

The sub-intensity matrix is given by

$$
\mathbf{S}_{[2:3]} = \begin{pmatrix}
\mathbf{G}_{(1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{G}_{(2)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{G}_{(3)} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_{(4)} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_{(5)} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_{(6)}
\end{pmatrix}
$$

where

$$
\mathbf{G}_{(1)} = -3\lambda_1, \quad \mathbf{G}_{(2)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1\\ 0 & -3\lambda_1 \end{pmatrix}, \quad \mathbf{G}_{(3)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0\\ 0 & -3\lambda_1 & 3\lambda_1\\ 0 & 0 & -3\lambda_1 \end{pmatrix}
$$

$$\mathbf{G}_{(4)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0 \\ 0 & -3\lambda_1 & 3\lambda_1 \\ 0 & 0 & -2\lambda_1 \end{pmatrix}, \quad \mathbf{G}_{(5)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0 & 0 \\ 0 & -3\lambda_1 & 3\lambda_1 & 0 \\ 0 & 0 & -3\lambda_1 & 3\lambda_1 \\ 0 & 0 & 0 & -2\lambda_1 \end{pmatrix},$$

$$\mathbf{G}_{(6)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0 & 0 \\ 0 & -3\lambda_1 & 3\lambda_1 & 0 \\ 0 & 0 & -2\lambda_1 & 2\lambda_1 \\ 0 & 0 & 0 & -2\lambda_1 \end{pmatrix}.$$

Lastly, a PH-representation for the distribution of $X_{[3:3]}$ is denoted by $\left(\boldsymbol{\alpha}_{[3:3]}, \mathbf{S}_{[3:3]}\right)$, where the vector of initial probabilities is given by

$$\boldsymbol{\alpha}_{(3:3)} = \left(\boldsymbol{\gamma}_{(1)}, \boldsymbol{\gamma}_{(2)}, \boldsymbol{\gamma}_{(3)}, \boldsymbol{\gamma}_{(4)}, \boldsymbol{\gamma}_{(5)}, \boldsymbol{\gamma}_{(6)}, \boldsymbol{\gamma}_{(7)}, \boldsymbol{\gamma}_{(8)}\right),$$

where

$$\boldsymbol{\gamma}_{(1)} = \frac{7}{54}, \quad \boldsymbol{\gamma}_{(2)} = \left(\frac{7}{54}, 0\right), \quad \boldsymbol{\gamma}_{(3)} = \left(\frac{2}{27}, 0, 0\right), \quad \boldsymbol{\gamma}_{(4)} = \left(\frac{1}{12}, 0, 0\right),$$

$$\boldsymbol{\gamma}_{(5)} = \left(\frac{1}{9}, 0, 0, 0\right), \quad \boldsymbol{\gamma}_{(6)} = \left(\frac{1}{12}, 0, 0, 0\right), \quad \boldsymbol{\gamma}_{(7)} = \left(\frac{2}{9}, 0, 0, 0, 0\right), \quad \boldsymbol{\gamma}_{(8)} = \left(\frac{1}{6}, 0, 0, 0, 0\right).$$

The sub-intensity matrix is given by

$$\mathbf{S}_{[3:3]} = \begin{pmatrix} \mathbf{H}_{(1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{(2)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_{(3)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{(4)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{(5)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{(6)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{(7)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{(8)} \end{pmatrix},$$

where

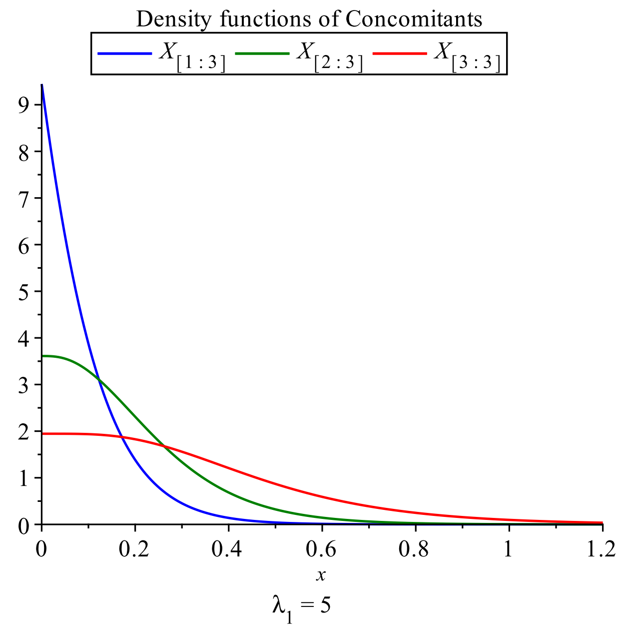
$$\mathbf{H}_{(1)} = -3\lambda_1, \quad \mathbf{H}_{(2)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 \\ 0 & -3\lambda_1 \end{pmatrix},$$

$$\mathbf{H}_{(3)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0 \\ 0 & -3\lambda_1 & 3\lambda_1 \\ 0 & 0 & -3\lambda_1 \end{pmatrix}, \quad \mathbf{H}_{(4)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0 \\ 0 & -3\lambda_1 & 3\lambda_1 \\ 0 & 0 & -2\lambda_1 \end{pmatrix},$$

$$\mathbf{H}_{(5)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0 & 0 \\ 0 & -3\lambda_1 & 3\lambda_1 & 0 \\ 0 & 0 & -3\lambda_1 & 3\lambda_1 \\ 0 & 0 & 0 & -2\lambda_1 \end{pmatrix}, \mathbf{H}_{(6)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0 & 0 \\ 0 & -3\lambda_1 & 3\lambda_1 & 0 \\ 0 & 0 & -2\lambda_1 & 2\lambda_1 \\ 0 & 0 & 0 & -2\lambda_1 \end{pmatrix},$$

$$\mathbf{H}_{(7)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0 & 0 & 0 \\ 0 & -3\lambda_1 & 3\lambda_1 & 0 & 0 \\ 0 & 0 & -3\lambda_1 & 3\lambda_1 & 0 \\ 0 & 0 & 0 & -2\lambda_1 & 2\lambda_1 \\ 0 & 0 & 0 & 0 & -\lambda \end{pmatrix},$$

$$\mathbf{H}_{(8)} = \begin{pmatrix} -3\lambda_1 & 3\lambda_1 & 0 & 0 & 0 \\ 0 & -3\lambda_1 & 3\lambda_1 & 0 & 0 \\ 0 & 0 & -2\lambda_1 & 2\lambda_1 & 0 \\ 0 & 0 & 0 & -2\lambda_1 & 2\lambda_1 \\ 0 & 0 & 0 & 0 & -\lambda \end{pmatrix}.$$



## 6.6 Concluding remarks

We have seen that concomitants of a finite sample of independent and identically bivariate phase-type distributions are phase-type distributed and also we have provided a procedure to calculate their densities.

The key of the results is essentially the application of the formula

$$\int_0^\infty e^{-ys} h(y|X \in dx) dy = \frac{\alpha(s)\, e^{\mathcal{W}(s)x} \eta(s)}{f_X(x)}$$

in Equation (6.35) and also that the function

$$\mathbb{E}\left(e^{-s\mathcal{Y}_2}, \mathcal{Y}_1 \in dy\right) = \alpha(s)\, e^{\mathcal{W}(s)y}\, \eta(s)\, dy \tag{6.48}$$

is analytic on the right-half plane. Therefore, the function (6.48) represents the link between the theory of stochastic fluid models and concomitants of phase-type distributions.

As a future work, we have missed to provide probabilistic arguments to generate PH-representations for the distribution of concomitants in the general case.

# APPENDIX A

# Formulas

**Lemma A.1 (The $n$-th derivative of the product of two vector functions.)**

$$\frac{d^n}{du^n}\left(\mathbf{p}(u)\cdot\mathbf{q}(u)\right) = \sum_{k=0}^{n}\binom{n}{k}\mathbf{p}(u)^{(k)}\cdot\mathbf{q}(u)^{(n-k)}, \tag{A.1}$$

*where $\mathbf{p}(u)^{(k)}$ means the $k$-th derivative of the vector $\mathbf{p}(u)$ and $\mathbf{p}(u)^{(0)} = \mathbf{p}(u)$,*

see [Bou04, pag. 20].

**Lemma A.2 (The $n$-th derivative of an inverse matrix with respect to a variable.)** *Consider a nonsingular matrix $\mathbf{A}$ in function of the variable $x$. Then*

$$\frac{d^n}{dx^n}\mathbf{A}^{-1} = n!\sum_{k=1}^{n}(-1)^k \sum_{\substack{1\le n_1,n_2,\dots,n_k\le n\\ n_1+\dots+n_k=n}} \left\{\mathbf{A}^{-1}\prod_{j=1}^{k}\left(\frac{1}{n_j!}\left(\frac{d^{n_j}}{dx^{n_j}}\mathbf{A}\right)\mathbf{A}^{-1}\right)_j\right\} \tag{A.2}$$

**Proof.** One proof is doable by an inductive procedure and by considering the first derivation of an inverse matrix, which is given by

$$\frac{d}{dx}\mathbf{A}^{-1} = -\mathbf{A}^{-1}\left(\frac{d}{dx}\mathbf{A}\right)\mathbf{A}^{-1}.$$

□

**Lemma A.3 (Integrals of a matrix exponential.)** *Let* $\mathbf{A}$ *be a nonsingular matrix. Then*

$$\int_0^\infty e^{\mathbf{A}t}dt = -\mathbf{A}^{-1}. \tag{A.3}$$

$$\int_0^\infty t^n e^{\mathbf{A}t}dt = n!\,(-\mathbf{A})^{-n}. \tag{A.4}$$

$$\int e^{\mathbf{A}t}dt = e^{\mathbf{A}t}\mathbf{A}^{-1}. \tag{A.5}$$

A proof is doable by induction.

**Lemma A.4 (Inverse of a partitioned matrix.)** *Let* $\mathbf{A} \in \mathbb{F}^{n\times n}, \mathbf{B} \in \mathbb{F}^{n\times m}, \mathbf{D} \in \mathbb{F}^{m\times m}$*, where the matrices* $\mathbf{A}$ *and* $\mathbf{D}$ *are nonsingular. Consider the block partitioned matrix given by*

$$\mathbf{Q} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}.$$

*Then, the inverse of* $\mathbf{Q}$ *is calculated in function of the sub-block matrices as follows*

$$\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{D}^{-1} \end{pmatrix}.$$

*See [Ber05, pag. 146].*

**Lemma A.5 (Power of an upper-block triangular matrix.)** *Let* $p, q \in \mathbb{N}$ *and let* $\mathbf{A} \in \mathbb{F}^{p\times p}, \mathbf{B} \in \mathbb{F}^{p\times q}$ *and* $\mathbf{C} \in \mathbb{F}^{q\times q}$*. Consider the upper-block partitioned matrix given by*

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0}_{q\times p} & \mathbf{C} \end{pmatrix},$$

*where* $\mathbf{0}_{q\times p}$ *is the matrix with zero-valued entries of dimension* $q \times p$*. Then, for every* $n \in \mathbb{N}$

$$\mathbf{\Lambda}^n = \begin{pmatrix} \mathbf{A}^n & \sum_{k=0}^{n-1} \mathbf{A}^{(n-1)-k}\mathbf{B}\mathbf{C}^k \\ \mathbf{0}_{q\times p} & \mathbf{C}^n \end{pmatrix}.$$

The proof is given by induction.

**Lemma A.6 (Matrix exponential of an infinitesimal generator of Phase-type.)** *Let* $X \sim PH(\boldsymbol{\alpha}, \mathbf{S})$ *and consider the infinitesimal generator of the underlying Markov jump process* $\mathbf{Q}$*. Then, the exponential of* $\mathbf{Q}$ *is given by*

$$e^{\mathbf{Q}t} = \begin{pmatrix} e^{\mathbf{S}t} & \mathbf{e} - e^{\mathbf{S}t}\mathbf{e} \\ \mathbf{0} & 1 \end{pmatrix},$$

*for all* $t \geq 0$.

**Proof.**

$$\begin{aligned}
e^{\mathbf{Q}t} &= \mathbf{I} + \sum_{n=1}^{\infty} \frac{t^n}{n!} \mathbf{Q}^n \\
&= \mathbf{I} + \sum_{n=1}^{\infty} \frac{t^n}{n!} \begin{pmatrix} \mathbf{S}^n & -\mathbf{S}^n \mathbf{e} \\ \mathbf{0} & 0 \end{pmatrix} \quad \text{(see Lemma A.5)} \\
&= \begin{pmatrix} \mathbf{I} + \sum\limits_{n=1}^{\infty} \frac{t^n}{n!} \mathbf{S}^n & -\sum_{n=1}^{\infty} \frac{t^n}{n!} \mathbf{S}^n \mathbf{e} \\ \mathbf{0} & 1 \end{pmatrix} \\
&= \begin{pmatrix} e^{\mathbf{S}t} & \mathbf{e} - e^{\mathbf{S}t} \mathbf{e} \\ \mathbf{0} & 1 \end{pmatrix}.
\end{aligned}$$

□

**Lemma A.7 (Kronecker product: the mixed-product property.)** *Let* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *and* $\mathbf{D}$ *be matrices with dimensions such that the products* $\mathbf{AB}$ *and* $\mathbf{CD}$ *are well-defined. Then*

$$(\mathbf{A} \otimes \mathbf{C}) (\mathbf{B} \otimes \mathbf{D}) = (\mathbf{AB}) \otimes (\mathbf{CD}). \tag{A.6}$$

**Faà de Bruno's formula: Generalisation of chain rule to higher derivatives.**

Let $n \in \mathbb{N}$. Then, the $n$-th derivative of the composite function $f(g(x))$ is calculated as

$$f^{(n)}(g(x)) = \sum \frac{n!}{m_1! \cdots m_n!} \left( f^{(m_1 + \cdots + m_2)}(g(x)) \right) \prod_{j=1}^{n} \left( \frac{g^{(j)}(x)}{j!} \right)^{m_j} \tag{A.7}$$

where the sum is over all the $n$-tuples of nonnegative integers $(m_1, m_2, \ldots, m_n)$ such that satisfy the equation

$$\sum_{j=1}^{n} j \, m_j = n.$$

# APPENDIX B

# List of symbols

In this appendix is explained the most common symbols and notation which is used throughout this thesis.

| Symbol | Description |
|---|---|
| $\otimes$ | Kronecker product. Let $\mathbf{A} = \{a_{i,j}\}$ and $\mathbf{B}$ be two rectangular matrices of dimension $k_1 \times k_2$ and $n_1 \times n_2$, respectively. Their Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the matrix of dimension $k_1 n_1 \times k_2 n_2$ given by $$\begin{pmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,k_2}\mathbf{B} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ a_{k_1,1}\mathbf{B} & a_{k_1,2}\mathbf{B} & \cdots & a_{k_1,k_2}\mathbf{B} \end{pmatrix}.$$ |
| $\oplus$ | Kronecker sum. Let $\mathbf{A} = \{a_{i,j}\}$ and $\mathbf{B}$ be two square matrices of dimension $k \times k$ and $n \times n$, respectively. The Kronecker sum of $\mathbf{A}$ and $\mathbf{B}$ is defined as $$\mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_k \otimes \mathbf{B}.$$ |
| $\bigoplus$ | Direct sum. Let $\mathbf{A}$ and $\mathbf{B}$ be any pair of matrices. Then the direct sum of $\mathbf{A}$ and $\mathbf{B}$ is defined as $$\mathbf{A} \bigoplus \mathbf{B} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}.$$ |

| | |
|---|---|
| $\mathbb{N}$ | Set of natural numbers. |
| $\mathbb{R}$ | Set of real numbers. |
| $\mathbb{C}$ | Set of complex numbers. |
| $\mathbb{N}_0$ | Set of natural numbers union with the number zero. |
| $\mathbb{N}_0^p$ | $\mathbb{N}_0 \times \cdots \times \mathbb{N}_0$, $p$ cartesian products of $\mathbb{N}_0$. |
| $\mathbb{P}$ | Probability measure. |
| $\mathbb{E}(X)$ | Expectation of the random variable $X$. |
| $\mathbf{1}_{\mathcal{A}}$ | Indicator function of the set $\mathcal{A}$. |
| $\mathbf{A}^n$ | $n$-power of the square matrix $\mathbf{A}$ and $n \in \mathbb{N}$. |
| $e^{\mathbf{A}}$ | Exponential of the square matrix $\mathbf{A}$. Formula: $e^{\mathbf{A}} = \sum\limits_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^n$. |
| $\rho(\mathbf{A})$ | Spectral radius of matrix $\mathbf{A}$. |
| $\mathbf{e}$ | Column vector of ones of appropriate dimension. |
| $\mathbf{e}_k$ | Column vector of 1-valued in the $k$-th entry and 0-valued in the rest of the entries. |
| $\mathbf{I}_q$ | Identity matrix of dimension $q$. |
| $\mathcal{R}e(s)$ | The real part value of the complex number $s$. |
| per$[\mathbf{A}]$ | Permanent of the square matrix $\mathbf{A}$. See definition in Equation (4.1). |
| det$(\mathbf{A})$ | Determinant of the matrix $\mathbf{A}$. |
| diag$(\mathbf{A}_i : i = 1, \ldots, n)$ | Block diagonal matrix of the square matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n$ (the direct sum of them). |
| $\lvert\mathcal{C}\rvert$ | If $\mathcal{C}$ is a set. Then $\lvert\mathcal{C}\rvert$ denotes the cardinality of $\mathcal{C}$. |
| $\lvert C\rvert$ | If $C$ is a number. Then $\lvert C\rvert$ denotes the absolute value of $C$. |

| | |
|---|---|
| $\chi(\mathbf{A})$ | The maximal real eigenvalue of the matrix $\mathbf{A}$. |
| $\sigma(\mathbf{A})$ | The spectrum of the matrix $\mathbf{A}$. |
| $\mathrm{PH}_p\left(\boldsymbol{\alpha}, \mathbf{S}\right)$ | Phase-type representation of dimension $p$. |
| $\mathrm{DPH}_p\left(\boldsymbol{\pi}, \mathbf{T}\right)$ | Discrete phase-type representaion of dimension $p$. |
| $\mathrm{ME}\left(\boldsymbol{\alpha}, \mathbf{S}, \mathbf{s}\right)$ | Matrix-exponential representation. |
| $\mathrm{MG}\left(\boldsymbol{\pi}, \mathbf{T}, \mathbf{t}\right)$ | Matrix-geometric representation. |
| $\mathrm{MPH}^*_p\left(\boldsymbol{\alpha}, \mathbf{S}\right)$ | Multivariate phase-type representation of dimension $p$. |
| $\mathrm{MDPH}^*_p\left(\boldsymbol{\pi}, \mathbf{T}\right)$ | Multivariate discrete phase-type representation of dimension $p$. |
| i.i.d. | Abbreviation for "independent and identically distributed". |
| $Y_{(r:n)}$ | $r$-th order statistic. |
| $Y_{[r:n]}$ | $r$-th concomitant. |
| $\mathcal{C}_{k,n}$ | Set of $k$-dimensional row vector formed with combinations of $k$ elements of the set $\{1, 2, \dots, n\}$. See Equation (4.4). |
| $F_X(x)$ | Cumulative distribution function of the random variable $X$. |
| $f_X(x)$ | Density function (or probability mass function) of the random variable $X$. |
| $\{X_t\}_{t\geq 0}$, $\{X_n\}_{n\in\mathbb{N}}$ | Markov jump process, Markov chain, respectively. |
| $\mathbf{P}$ | Matrix of transition probabilities of. |
| $\mathcal{E}$ | State space of the Markov jump process or Markov chain. |
| $\binom{n}{i}$ | For $n, i \in \mathbb{N}$, where $n \geq i$, it denotes the number of combinations formed with $i$ elements from a set of $n$ elements. |
| $< \cdot, \cdot >$ | Dot product. |

# Bibliography


[AB95] S. Asmussen and M. Bladt. Renewal Theory and Queueing Algorithms for Matrix-Exponential distributions. In A. S. Alfa and S. R. Chakravarthy, editors, *Matrix-analytic methods in stochastic models*, volume 183, pages 313–341. Marcel Dekker, New York, 1995.

[Abd11] Y. H. Abdelkader. A Laplace transform method for order statistics from nonidentical random variables and its application in Phase-type distribution. *Statistics & Probability Letters*, 81(8):1143–1149, 2011.

[ABN92] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. SIAM, 1992.

[ALSS84] D. Assaf, N. Langberg, T. Savits, and M. Shaked. Multivariate phase-type distributions. *Operations Research*, 32(3):688–702, 1984.

[BA04] H. M. Barakat and Y. H. Abdelkader. Computing the moments of order statistics from nonidentical random variables. *Statistical Methods & Applications*, 13(1):15–26, 2004.

[Bal07] N. Balakrishnan. *Permanents, order statistics, outliers, and robustness*, volume 20. Revista matemática complutense, 2007.

[BB89] R. B. Bapat and M. I. Beg. Order Statistics for non-identically distributed variables and permanents. *The Indian Journal of Statistics*, 51(1):79–93, 1989.

[BB95] K. Balasubramanian and N. Balakrishnan. On a class of multivariate distributions closed under concomitance of order statistics. *Statistics and Probability Letters*, 23(3):239–242, 1995.

[Ber05] D. Bernstein. *Matrix Mathematics*. Princeton University Press, 2005.

[BH11] J. Berdel and C. Hipp. Convolutions of multivariate phase-type distributions. *Insurance: Mathematics and Economics*, 48(3):374–377, 2011.

[Bha84] P. K. Bhattacharya. Induced order statistics: Theory and Applications. *Krishnaiah, P. R. Sen. P.K. (Eds.), Handbook of Statistics. In: Nonparametric Methods, Elsevier, Amsterdam*, 4:383–403, 1984.

[BKJ97] N. Balakrishnan, S. Kotz, and N. L. Johnson. *Discrete Multivariate Distributions*. Wiley Series in Probability and Statistics, 1997.

[BN17] M. Bladt and B. F. Nielsen. *Matrix-Exponential Distributions in Applied Probability*, volume 81. Probability Theory and Stochastic Modelling, Springer, 2017.

[BO13] N.G. Bean and M. M. O'Reilly. A stochastic two-dimensional fluid model. *Stochastic Models*, 29:31–63, 2013.

[BOT05] N. G. Bean, M.M. O'Reilly, and P.G. Taylor. Hitting probabilities and hitting times for stochastic fluid flows. *Stochastic Processes and their Applications*, 115:1530–1556, 2005.

[Bou04] N. Bourbaki. *Functions of a real variable: elementary theory*. Springer, 2004.

[Bre12] L. Breuer. Occupation times for Markov-modulated Brownian motion. *Journal of Applied Probability*, 49(2):549–565, 2012.

[Bre16] L. Breuer. A semi-explicit density function for Kulkarni's bivariate phase-type distribution. *Stochastic Models*, 32(4):632–642, 2016.

[Cox55] D. R. Cox. A use of complex probabilities in the theory of stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(2):313–319, 1955.

[Dav73] H. A. David. Concomitants of order statistics. *Bull. Internat. Statist.*, 45:295–300, 1973.

[DN98] H. A. David and H. N. Nagaraja. Concomitants of Order Statistics. *N. Balakrishnan and C. R. Rao, eds., Handbook of Statistics, Elsevier Science B.V.*, 16:487–513, 1998.

[DN03] H. A. David and H. N. Nagaraja. *Order Statistics*. New York: John Wiley and Sons. Third edition., 2003.

[DN04] H. A. David and H. N. Nagaraja. *Order Statistics*. John Wiley & Sons, April 2004.

[Dun67] J. E. Dunn. Characterization of the Bivariate Negative Binomial Distribution. *Journal of the Arkansas Academy of Science*, 21(17), 1967.

[EG61] C. B. Edwards and J. Gurland. A Class of Distributions Applicable to Accidents. *Journal of the American Statistical Association*, 56(295):503–517, 1961.

[Erl17] A.K. Erlang. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Post Office Electrical Engineer's Journal*, 10:189–197, 1917.

[Ery05] S. Eryilmaz. Concomitants in a sequence of independent nonidentically distributed random vector. *Communications in Statistics–Theory and Methods*, 34:1925–1933, 2005.

[Fac03] M. W. Fackrell. *Characterization of matrix-exponential distributions*. PhD thesis, School of Applied Mathematics, University of Adelaide, 2003.

[Gau97] N. Gautam. *Quality of service of multi-class traffic in high-speed networks*. PhD thesis, Department of Operations Research, University of North Carolina, Chapel Hill, 1997.

[Gre09] S. Greeuw. *On the relation between Matrix-Geometric and discrete Phase-type distributions*. Master's Thesis, University of Amsterdam, 2009.

[HA17] Q.-M. He and A. S. Alfa. Space reduction for a class of multidimensional Markov chains: A summary and some applications. *Informs Journal on Computing*, 30(1):1–10, 2017.

[Hog07] L. Hogben. *Handbook of Linear Algebra*. Chapman & Hall/CRC, Taylor & Francis Group, 2007.

[HR16a] Q.-M. He and J. Ren. Analysis of a multivariate claim process. *Methodology and Computing in Applied Probability*, 18(1):257–273, 2016.

[HR16b] Q.-M. He and J. Ren. Parameter estimation of discrete multivariate phase-type distributions. *Methodology and Computing in Applied Probability*, 18(3):629–651, 2016.

[HS79] F. E. Harrell and P. K. Sen. Statistical Inference for Censored Bivariate Normal Distributions based on Induced Order Statistics. *Biometrika*, 66(2):293–298, 1979.

[Jen53] A. Jensen. *A distribution model applicable to Economics*. Munksgaard, Copenhagen, 1953.

[Kam95] U. Kamps. A concept of generalized order statistics. *Journal of Statistical Planning and Inference*, 48(1):1–23, 1995.

[Kru98] F. Krummenauer. Representation of multivariate discrete distributions by probability generating functions. *Statistics & Probability Letters*, 39(4):327–331, 1998.

[Kul89] V. G. Kulkarni. A new class of Multivariate Phase type distributions. *Operations Research*, 37(1):151–158, 1989.

[Kum08] C. S. Kumar. A unified approach to bivariate discrete distributions. *Metrika*, 67(1):113–123, 2008.

[LR99] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modelling*. ASA-SIAM, Series on Statistics and Applied Mathematics, 1999.

[LRS+12] G. Latouche, V. Ramaswami, J. Sethuraman, K. Sigman, M. Squillante, and D. Yao. *Matrix-Analytic Methods in Stochactic models*. Springer, 2012.

[MNOS10] C. Min-Ng, S.-H. Ong, and H. M. Srivastava. A class of bivariate negative binomial distribution with different index parameters in the marginals. *Applied Mathematics and Computation*, 217(7):3069–3087, 2010.

[Neu75] M. F. Neuts. *Probability distributions of phase-type*, pages 173–206. In *Liber Amicorum Prof. Emeritus H. Florin*, Department of Mathematics. Belgium: University of Louvain, 1975.

[Neu81] M. F. Neuts. *Matrix-geometric Solutions in Stochastic Models*. An Algorithmic Approach. Courier Dover Publications, 1981.

[OC90a] C. A. O' Cinneide. On the limitations of multivariate phase-type families. *Operations Research*, 38(3):519–526, 1990.

[O'C90b] C. A. O'Cinneide. Characterization of phase-type distributions. *Commun. Statist. - Stochastic models*, 6(1):1–57, 1990.

[Rén53] A. Rényi. On the theory of order statistics. *Acta Mathematica Hungarica*, 4(3):191–231, 1953.

[Rui13] J. Ruiz. Matrix-exponential distributions: closure properties. *International Journal of Advanced Statistics and Probability*, 1(2):44–52, 2013.

[RZ17] J. Ren and R. Zitikis. CMPH: a multivariate phase-type aggregate loss distribution. *Dependence Modeling*, 5(1):304–315, 2017.

[Ste98] G. W. Stewart. *Matrix Algorithms. Volume 1: Basic Decompositions*. SIAM, 1998.

[TV17] P. Y. Thomas and T. G. Veena. Role of concomitants of order statistics in determining parent bivariate distributions. *Communications in Statistics–Theory and Methods*, 46(16):7976–7997, 2017.

[VL78] C. F. Van Loan. Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control*, 23(3):395–404, 1978.

[WN09] K. Wang and H. N. Nagaraja. Concomitants of order statistics for dependent samples. *Statistics and Probability Letters*, 79(4):553–558, 2009.

[Yan77] S. S. Yang. General distribution theory of concomitants of order statistics. *The Annals of Statistics*, 5(5):996–1002, 1977.

[ZH11] X. Zhang and Z. Hou. Computing the moments of order statistics from nonidentically distributed phase-type random variables. *Journal of Computational and Applied Mathematics*, 235(9):2897–2903, 2011.